To climb steep hills requires slow pace at first.

William Shakespeare, English
poet and playwright
(1564 - 1818)

Don't buy the house; buy the neighborhood.

Russian Proverb

2 Perturbation Analysis

Perturbation analysis (PA) is the core of the gradient-based (or policy gradient) learning and optimization approach. The basic principle of PA is that *the derivative of a system's performance with respect to a parameter of the system can be decomposed into the sum of many small building blocks, each of which measures the effect of a single perturbation on the system's performance, and this effect can be estimated on a sample path of the system.* This decomposition principle applies to the differences in a system's performance with two policies as well and is thus fundamental to other learning and optimization approaches such as the policy iteration approach (see Chapter 4).

Historically, perturbation analysis was first developed for queueing systems and was later extended to Markov systems. Because PA of Markov systems is generally applicable and has a strong connection with other learning and optimization approaches, such as Markov decision processes and reinforcement learning, we first introduce the PA principle to Markov systems. PA of queueing systems will be discussed at the end of this chapter as supplementary material.

There were a number of books published in later 1980's and 1990's on PA of queueing-type systems [51, 72, 107, 112, 142]. The PA principle summarized above was discussed in detail in [45, 51, 141, 142] and extended to Markov systems in [62, 70].

2.1 Perturbation Analysis of Markov Chains

We first discuss PA of discrete-time Markov chains and related topics in this section. PA of continuous-time Markov processes is covered in the next section.

Consider an ergodic (irreducible and aperiodic) Markov chain $\boldsymbol{X} = \{X_l : l \geq 0\}$ on a finite state space $\mathcal{S} = \{1, 2, \ldots, S\}$ with a transition probability matrix $P = [p(j|i)]_{i,j=1}^{S}$. Its steady-state probabilities are denoted as a row vector $\pi = (\pi(1), \ldots, \pi(S))$ and the reward function is denoted as a (column) vector $f = (f(1), f(2), \ldots, f(S))^T$. We have $Pe = e$, where $e = (1, 1, \ldots, 1)^T$, and the probability flow balance equation $\pi = \pi P$. We first consider the *long-run average reward* (or, simply, the average reward) as the performance measure, which is defined as follows:

$$\eta = E_\pi(f) = \sum_{i=1}^{S} \pi(i) f(i) = \pi f,$$

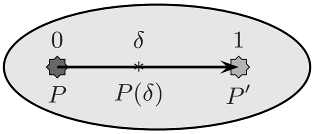
where E_π denotes the expectation corresponding to the steady-state probability π on $\mathcal{S}$.

Let P' be another irreducible and aperiodic transition probability matrix on the same state space $\mathcal{S}$. Suppose that P changes to

$$P_\delta = P + \delta \Delta P = \delta P' + (1 - \delta) P, \tag{2.1}$$

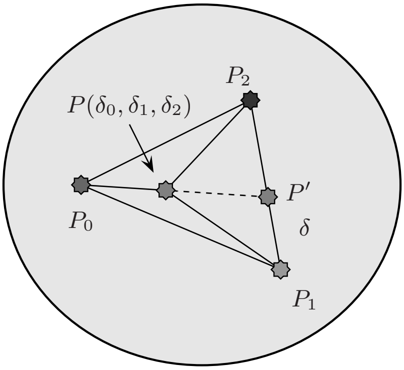
with $0 \leq \delta \leq 1$ and $\Delta P = P' - P := [\Delta p(j|i)]$. Since $Pe = P'e = e$, we have $(\Delta P)e = 0$ and $P_\delta e = e$.

P_δ represents a randomized policy, which, at every state transition, implements policy P with probability $1 - \delta$ and policy P' with probability δ. When δ varies from 0 to 1, P_δ fills the line from P to P' in the policy space (Figure 2.1). With randomized policies, we can fill all the policies in the convex set spanned by a set of policies in a policy space. For example, we can fill the triangle with vertices P_0, P_1, and P_2 in the policy space with randomized policies $P(\delta_0, \delta_1, \delta_2) := \delta_0 P_0 + \delta_1 P_1 + \delta_2 P_2$, where $\delta_0 + \delta_1 + \delta_2 = 1$; $P(\delta_0, \delta_1, \delta_2)$ implements policy P_i with probability δ_i, $i = 0, 1, 2$, at every state transition (Figure 2.2).



$$P(\delta) = \delta P' + (1 - \delta) P$$
$$P(0) = P, P(1) = P'$$

Fig. 2.1. Randomized Policies with Two Base Policies



$$P(\delta_0, \delta_1, \delta_2) = \delta_0 P_0 + \delta_1 P_1 + \delta_2 P_2$$
$$P' = P(0, 1-\delta, \delta)$$
$$P_0 = P(1,0,0), P_1 = P(0,1,0), P_2 = P(0,0,1)$$

Fig. 2.2. Randomized Policies with Three Base Policies

For simplicity, we first assume that the Markov chain with transition probability matrix P_δ in (2.1) for all $0 \le \delta \le 1$ has the same reward function f, and we denote it as (P_δ, f). The steady-state probability of transition matrix P_δ is denoted as π_δ and the average reward of the Markov chain (P_δ, f) is denoted as $\eta_\delta = \pi_\delta f$. Then $\eta_0 = \eta = \pi f$ and $\eta_1 = \eta' = \pi' f$. Set $\Delta\eta_\delta = \eta_\delta - \eta$. The derivative of η_δ with respect to δ at $\delta = 0$ is

$$\left.\frac{d\eta_\delta}{d\delta}\right|_{\delta=0} = \lim_{\delta\to 0} \frac{\Delta\eta_\delta}{\delta},$$

which can be viewed as the directional derivative in the policy space along the direction from policy P to policy P' (see Figure 1.9 and Figure 2.1).

The goal of perturbation analysis is to determine the performance derivative $\frac{d\eta_\delta}{d\delta}$ by observing and/or analyzing the behavior of the Markov chain with transition probability matrix P. In particular, we wish to estimate this derivative by observing and analyzing a single sample path of the Markov chain with transition probability matrix P.

2.1.1 Constructing a Perturbed Sample Path

The main idea of PA comes from the fact that given a sample path of the Markov chain with transition probability matrix P, we can construct a sample path of the Markov chain with transition probability matrix P_δ, when δ is small; and this does not require that we rerun or resimulate the Markov chain with P_δ. If δ is small, the additional computation involved is also small. The performance derivative $\frac{d\eta_\delta}{d\delta}$ can be obtained by measurement or analysis once we have the sample paths of both P and P_δ. The above statement as well as the construction procedure described below is not very precise, but they provide a clear intuition and help us to derive the performance derivative formula, which will be proved rigorously later.

Following the PA terminology, we call the Markov chain with transition probability matrix P the *original Markov chain*, and that with P_δ the *perturbed Markov chain*. Their sample paths are called the *original sample paths* and the *perturbed sample paths*, respectively.

Constructing a Sample Path

We first review how to simulate a sample path for a Markov chain with transition probability matrix P. Suppose that at time $l = 0, 1, \dots,$ the Markov chain is in state $X_l = k$. In simulation, the next state after the transition at any time l is determined as follows. We generate a uniformly distributed random variable $\xi_l \in [0, 1)$. If

$$\sum_{k'=1}^{u_{l+1}-1} p(k'|k) \le \xi_l < \sum_{k'=1}^{u_{l+1}} p(k'|k), \qquad u_{l+1} \in \mathcal{S}, \tag{2.2}$$

(with the convention $\sum_{k'=1}^{0} p(k'|k) = 0$), then we set $X_{l+1} = u_{l+1}$. In the case illustrated in Figure 2.3.A, we have $p(i|k) = 0.5$, $p(j|k) = 0.5$, and $p(k'|k) = 0$ for all $k' \neq i, j$. The current state is k. We generate a $[0, 1)$-uniformly distributed random variable ξ. If $0 \le \xi < 0.5$, then the Markov chain moves into state i; otherwise, it moves into state j. Following this process, starting from any initial state X_0, we can construct a sample path for the Markov chain with any transition probability matrix P. Therefore, a sample path of a Markov chain is determined by an initial state X_0 and a sequence of $[0, 1)$-uniformly distributed random variables $\{\xi_0, \xi_1, \dots\}$. Figure 2.4 illustrates such a sample path $\boldsymbol{X} := \{X_0, X_1, \dots, X_l, \dots\}$.

The performance measure η can be estimated from the sample path $\boldsymbol{X}$. In fact, if the Markov chain is ergodic, we have

$$\eta = \lim_{L\to\infty} \frac{1}{L} \sum_{l=0}^{L-1} f(X_l), \qquad \text{w.p.1},$$

where “w.p.1” stands for “with probability 1”. Set

$$F_L = \sum_{l=0}^{L-1} f(X_l).$$

Then, we have

$$\eta = \lim_{L\to\infty} \frac{F_L}{L}. \tag{2.3}$$

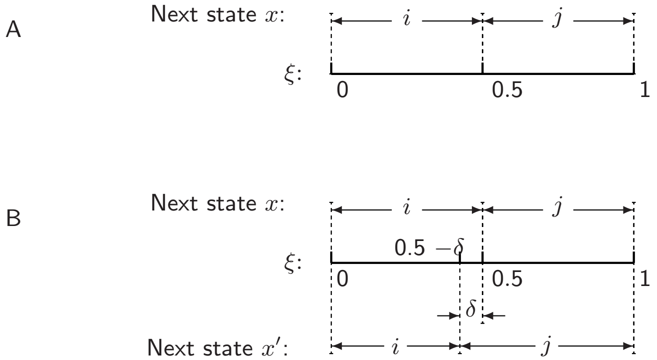


Fig. 2.3. Determining the State Transitions

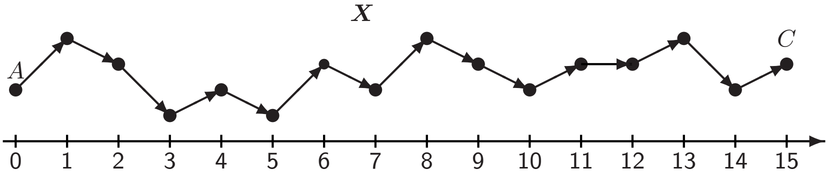


Fig. 2.4. A Sample Path of a Markov Chain

Constructing a Perturbed Sample Path on a Given Original Sample Path

Now, suppose that we are given a sample path of a Markov chain with transition probability matrix P, as shown as $\boldsymbol{X} = \{X_0, X_1, \dots\}$ in Figure 2.4. It starts with initial state X_0 and is generated according to (2.2) with a sequence of $[0,1)$-uniformly distributed and independent random numbers $\{\xi_0, \xi_1, \dots, \xi_l, \dots\}$. We wish to construct a perturbed sample path for the Markov chain with $P_\delta = P + \delta \Delta P$. We denote it as $\boldsymbol{X}_\delta = \{X_{\delta,0}, X_{\delta,1}, \dots\}$. To this end, we may think as follows.

To save computation, we may try to use the same sequence $\{\xi_0, \xi_1, \dots, \xi_l, \dots\}$ to generate the perturbed path. However, we need to use (cf. (2.2))

$$\sum_{k'=1}^{u_{\delta,l+1}-1} [p(k'|k) + \delta \Delta p(k'|k)] \le \xi_l < \sum_{k'=1}^{u_{\delta,l+1}} [p(k'|k) + \delta \Delta p(k'|k)] \qquad (2.4)$$

to determine the state at $X_{\delta,l+1}$; i.e., if (2.4) holds, we set $X_{\delta,l+1} = u_{\delta,l+1}$.

First, we observe that when δ is very small, in most cases we may have $u_{\delta,l+1} = u_{l+1}$, if $X_{\delta,l} = X_l$, $l = 0, 1, \dots$. For example, let us assume that the transition probabilities of the Markov chain in Figure 2.3 are perturbed to $p_\delta(i|k) = 0.5-\delta$, $p_\delta(j|k) = 0.5+\delta$, and $p_\delta(l|k) = 0$, $l \neq i, j$ (i.e., $\Delta p(i|k) = -1$, $\Delta p(j|k) = 1$, and $\Delta p(l|k) = 0$). In this case, if $X_{\delta,l} = X_l = k$ and the same ξ_l is used to determine the state transition, then $X_{\delta,l+1} \neq X_{l+1}$ if and only if $0.5 - \delta \leq \xi_l < 0.5$, in which case the original Markov chain $\boldsymbol{X}$ moves to $X_{l+1} = i$, but the perturbed one $\boldsymbol{X}_\delta$ moves to $X_{\delta,l+1} = j$. The probability that this discrepancy occurs is δ, which is very small as assumed, see Figure 2.3.B.

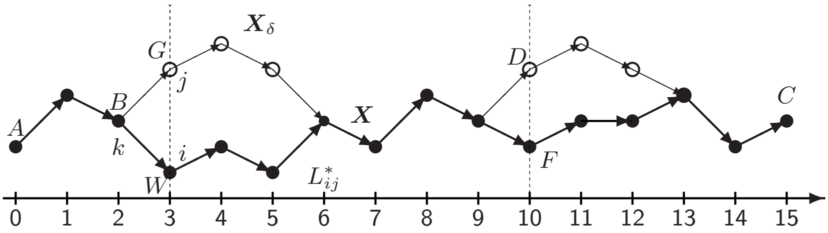


Fig. 2.5. Constructing a Perturbed Sample Path

Now we start with the same initial state $X_0 = X_{\delta,0}$ to construct the perturbed path. This procedure is illustrated in Figure 2.5, in which point A denotes the initial state and path $A-B-W-F-C$ is the given original sample path $\boldsymbol{X}$. As we have explained, starting from the same state the transitions of $\boldsymbol{X}$ and $\boldsymbol{X}_\delta$ differ only with a very small probability. In Figure 2.5, it so happens that with the same random variables ξ_0 and ξ_1, according to (2.2) and (2.4), we have $X_1 = X_{\delta,1}$ and $X_2 = X_{\delta,2}$.

Next, we assume that at $l = 2$, according to (2.2) and (2.4) with the same random variable ξ_2, we determine that $\boldsymbol{X}$ moves to $X_3 = i$ (point W) but $\boldsymbol{X}_\delta$ moves to another state $X_{\delta,3} = j$ (point G). We say that, because of the change of P to P_δ, the system has a *perturbation* (or simply called a "jump") from i to j at $l = 3$. After $l = 3$, the original sample path follows the path $W - F - C$; the perturbed path, however, follows a completely different path starting from point G. For convenience in understanding, let us generate an additional sequence of $[0, 1)$-uniformly and independently distributed random variables $\xi_{\delta,3}, \xi_{\delta,4}, \ldots$, which are also independent of $\xi_3, \xi_4, \ldots$, to construct the perturbed path following (2.4) starting from point G at $l = 3$ until the perturbed path merges with the original one.

Figure 2.5 shows that the perturbed path $\boldsymbol{X}_\delta$ merges with the original one $\boldsymbol{X}$ at $l = L^*_{ij} = 6$. Theoretically, because both sample paths $\boldsymbol{X}$ and $\boldsymbol{X}_\delta$ are ergodic, they will merge in finite steps (i.e., L^*_{ij} is finite) with probability 1. Let $\xi_{\delta,3}, \xi_{\delta,4}$, and $\xi_{\delta,5}$ be the random variables that determine the transitions at $l = 3, 4$, and 5 (or equivalently, the states $X_{\delta,4}, X_{\delta,5}$, and $X_{\delta,6} = X_6$) on $\boldsymbol{X}_\delta$. Then, the original path $\boldsymbol{X}$ from X_0 to X_6 is generated by $\xi_0, \xi_1, \xi_2, \xi_3, \xi_4, \xi_5$, while the perturbed path $\boldsymbol{X}_\delta$ from $X_{\delta,0}$ to $X_{\delta,6}$ is generated by $\xi_0, \xi_1, \xi_2, \xi_{\delta,3}, \xi_{\delta,4}, \xi_{\delta,5}$, with $\xi_{\delta,3}, \xi_{\delta,4}, \xi_{\delta,5}$ independent of ξ_3, ξ_4, ξ_5.

Starting from the merging point $X_6 = X_{\delta,6}$, the situation is the same as at the initial point $X_0 = X_{\delta,0}$. Again, we use the same random variables $\xi_6, \xi_7, \ldots$, to construct the perturbed path until it differs from the original one. In Figure 2.5, it so happens that with the same random variables ξ_6, ξ_7, and ξ_8 according to (2.2) and (2.4) we have $X_7 = X_{\delta,7}$, $X_8 = X_{\delta,8}$ and $X_9 = X_{\delta,9}$. However, there is a perturbation at $l = 10$. In other words, according to (2.2) and (2.4) with the same random variable ξ_9, we determine that $\boldsymbol{X}$ and $\boldsymbol{X}_\delta$ move to two different states X_{10} (point F) and $X_{\delta,10}$ (point D), respectively. After $l = 10$, the situation is the same as at $l = 3$. The two sample paths $\boldsymbol{X}$ and $\boldsymbol{X}_\delta$ follow different paths $D - C$ and $F - C$ until they merge again at $l = 13$. $X_{\delta,11}$, $X_{\delta,12}$, and $X_{\delta,13}$ are generated by random variables $\xi_{\delta,10}$, $\xi_{\delta,11}$, and $\xi_{\delta,12}$, which are independent of ξ_{10}, ξ_{11}, and ξ_{12}. $\boldsymbol{X}_\delta$ and $\boldsymbol{X}$ merge again at $l = 13$. Starting from this merging point $X_{\delta,13} = X_{13}$, once again the situation is the same as at the initial point $X_0 = X_{\delta,0}$.

The above description illustrates how to construct a perturbed sample path $\boldsymbol{X}_\delta$, given an original sample path $\boldsymbol{X}$. At any time instant l, if $X_{\delta,l} = X_l$, then we use the same random variable ξ_l to determine the state transitions (or equivalently X_{l+1} and $X_{\delta,l+1}$) for both $\boldsymbol{X}$ and $\boldsymbol{X}_\delta$ by using (2.2) and (2.4); if it turns out that $X_{\delta,l+1} \neq X_{l+1}$, we say there is a perturbation (jump) at $l + 1$. After each jump, $\boldsymbol{X}_\delta$ is completely different from $\boldsymbol{X}$ until they merge together. In these segments in which the two sample paths are different, $\boldsymbol{X}_\delta$ is generated independently of $\boldsymbol{X}$. In Figure 2.5, $\boldsymbol{X}$ and $\boldsymbol{X}_\delta$ are generated by the following sequences of random variables, respectively:

$$\begin{array}{l}
\boldsymbol{X} : \xi_0 \; \xi_1 \; \xi_2 \;\; \xi_3 \;\;\; \xi_4 \;\;\; \xi_5 \;\; \xi_6 \; \xi_7 \; \xi_8 \; \xi_9 \;\; \xi_{10} \;\;\; \xi_{11} \;\;\; \xi_{12} \;\; \xi_{13} \; \xi_{14}, \\
\boldsymbol{X}_\delta : \xi_0 \; \xi_1 \; \xi_2 \; \xi_{\delta,3} \; \xi_{\delta,4} \; \xi_{\delta,5} \; \xi_6 \; \xi_7 \; \xi_8 \; \xi_9 \; \xi_{\delta,10} \; \xi_{\delta,11} \; \xi_{\delta,12} \; \xi_{13} \; \xi_{14},
\end{array}$$

where all the random variables ξ_l, and $\xi_{\delta,l}$ are independent of each other.

Finally, when δ is very small, perturbations rarely happen (see Figure 2.3.B). In this case, in most time instants, the perturbed sample path $\boldsymbol{X}_\delta$ is the same as the original one $\boldsymbol{X}$; i.e., in reality, the lengths of the common segments are much longer than what might be indicated by $X_0 - X_2$, $X_6 - X_9$, and $X_{13} - X_{14}$ in Figure 2.5.

2.1.2 Perturbation Realization Factors and Performance Potentials

To calculate performance derivatives, we need to compare the average rewards of the original and the perturbed Markov chains, η and η_δ, by using the sample paths $\boldsymbol{X}$ and $\boldsymbol{X}_\delta$ constructed above. As shown in Figure 2.5, the difference between $\boldsymbol{X}$ and $\boldsymbol{X}_\delta$ is only reflected in the segments after the perturbations. In other words, the effect of a change of the transition probability matrix from P to P_δ on the system performance can be decomposed into the sum of the effects of the perturbations generated due to the change in P. Therefore, we first need to study the effect of a single perturbation on the system performance. We show that this effect can be measured by a quantity called the *perturbation realization factor.*

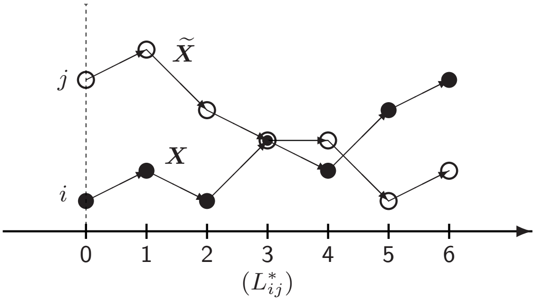


Fig. 2.6. Realization of a Perturbation

Perturbation Realization

Again, we use Figure 2.5 to illustrate the idea. At $l = 3$, the sample path is perturbed from state i to state j. This perturbation will certainly affect the system's behavior and the system's performance. As shown in Figure 2.5, after $l = 3$, the perturbed Markov chain evolves differently from the original chain, until, at $l = L^*_{ij}$, the perturbed path merges with the original one. The effect of the perturbation takes place in the period from $l = 3$ to L^*_{ij}. In PA terminology, we say that the perturbation generated at $l = 3$ is *realized* by the system at $l = L^*_{ij} = 6$.

Strictly speaking, the perturbed path $\boldsymbol{X}_\delta$ follows the perturbed transition probability matrix P_δ. However, because δ is very small and the length from the perturbed point $l = 3$ to the merging point L^*_{ij}, $L^*_{ij} - 3$, is finite (with probability 1), the probability that there is another perturbation in the period from $l = 3$ to L^*_{ij} (i.e., there are two perturbations in the period from $l = 3$ to L^*_{ij}, one at $l = 3$ the other in the period from $l = 4$ to L^*_{ij}) is on the order δ^2. This contributes to the high-order performance derivatives and in the first-order derivatives we may ignore this high-order term. Therefore, to calculate the performance derivatives, as δ approaches zero, we may assume that from $l = 3$ to L^*_{ij} the perturbed path $\boldsymbol{X}_\delta$ is the same as if it follows the original transition probability matrix P.

Thus, to quantify the effect of a single perturbation from i to j, we study two independent Markov chains $\boldsymbol{X} = \{X_l, l \geq 0\}$ and $\widetilde{\boldsymbol{X}} = \{\widetilde{X}_l, l \geq 0\}$ with $X_0 = i$ and $\widetilde{X}_0 = j$, respectively; both of them follow the same transition matrix P (Figure 2.6). Let these two sample paths merge for the first time at L^*_{ij}, i.e.,

$$\boxed{L^*_{ij} = \min\left\{ l : l \geq 0, \widetilde{X}_l = X_l \middle| \widetilde{X}_0 = j, X_0 = i \right\}.}$$

Recall that the performance measure is $\eta \approx \frac{F_L}{L}$ (see (2.3)). Apparently, the average effect of a single perturbation on η is zero, because L^*_{ij} is finite with probability 1. We, therefore, study the effect of a single perturbation on F_L for a large L.

Let E denote the expectation in the probability space spanned by all the sample paths of both $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}$. The *perturbation realization factor (PRF)* is defined as [62, 70]:

$$\boxed{\gamma(i,j) = E\left\{ \sum_{l=0}^{L^*_{ij}-1} \left[f(\widetilde{X}_l) - f(X_l) \right] \middle| \widetilde{X}_0 = j,\ X_0 = i \right\}, \quad i,j = 1,\ldots,S. \quad (2.5)}$$

Thus, $\gamma(i,j)$ represents the average effect of a jump from i to j on F_L in (2.3). For convenience, sometimes we may refer to $\gamma(i,j)$ as the effect of a jump on the performance η itself, although this effect is on an "infinitesimal" scale.

By the strong Markov property, the two Markov chains $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}$ behave similarly statistically after L^*_{ij}. Thus,

$$\lim_{L\to\infty} E\left\{ \sum_{l=L^*_{ij}}^{L-1} \left[f(\widetilde{X}_l) - f(X_l) \right] \middle| \widetilde{X}_0 = j,\ X_0 = i \right\} = 0.$$

Therefore, (2.5) becomes

$$\begin{aligned} \gamma(i,j) &= \lim_{L\to\infty} E\left\{ \sum_{l=0}^{L-1} \left[f(\widetilde{X}_l) - f(X_l) \right] \middle| \widetilde{X}_0 = j,\ X_0 = i \right\} \\ &= \lim_{L\to\infty} E\left[\widetilde{F}_L - F_L \middle| \widetilde{X}_0 = j,\ X_0 = i, \right], \qquad i,j = 1,\ldots,S. \quad (2.6) \end{aligned}$$

Essentially, the perturbation realization factors use the difference in the sums of the rewards on the perturbed path and the original one to measure the effect of a single perturbation.

The matrix $\Gamma := [\gamma(i,j)]_{i,j=1}^S \in \mathcal{R}^{S\times S}$ is called a *perturbation realization factor (PRF) matrix*. From (2.5), we have

$$\gamma(i,j) = f(j) - f(i) + \sum_{i'=1}^{S}\sum_{j'=1}^{S} E\left\{ \sum_{l=1}^{L^*_{i'j'}-1} \left[f(\widetilde{X}_l) - f(X_l)\right] \middle| \widetilde{X}_1 = j', X_1 = i' \right\}$$
$$\times \mathcal{P}\left(\widetilde{X}_1 = j', X_1 = i' \middle| \widetilde{X}_0 = j, X_0 = i\right)$$
$$= f(j) - f(i) + \sum_{i'=1}^{S}\sum_{j'=1}^{S} p(i'|i)p(j'|j)\gamma(i',j').$$

By writing this in a matrix form, we have the following *PRF equation* [70]

$$\Gamma - P\Gamma P^T = F, \qquad (2.7)$$

where $F = ef^T - fe^T$.

If F is a Hermitian matrix, then (2.7) is called the Lyapunov equation in the literature [13, 14, 162, 174]. (A *Hermitian matrix*, also called a self-adjoint matrix, is a square matrix that is equal to its own conjugate transpose. Thus, a real Hermitian matrix is a symmetric matrix.) The PRF equation differs from the Lyapunov equation because F is a skew-symmetric matrix, $F^T = -F$.

Performance Potentials

From (2.6), we have $\gamma(i,i) = 0$ for any $i = 1, \ldots, S$, and $\gamma(i,j) = -\gamma(j,i)$, or $\Gamma^T = -\Gamma$; i.e., Γ is skew-symmetric. In addition, from (2.6), we can easily prove

$$\gamma(i,j) = \gamma(i,k) + \gamma(k,j), \qquad i,j,k = 1, \ldots, S. \qquad (2.8)$$

This is the same equation as that for the differences of potential energies in physics. This observation motivates the following analysis: Let us fix any state denoted as $k^* \in \mathcal{S}$. Then, (2.8) becomes

$$\gamma(i,j) = \gamma(i,k^*) + \gamma(k^*,j) = \gamma(k^*,j) - \gamma(k^*,i), \qquad i,j = 1, \ldots, S.$$

Define $g_{k^*}(j) = \gamma(k^*,j)$. Then,

$$\gamma(i,j) = g_{k^*}(j) - g_{k^*}(i), \qquad i,j = 1, \ldots, S. \qquad (2.9)$$

For any two states k_1^* and k_2^*, we have

$$g_{k_2^*}(j) - g_{k_1^*}(j) = \gamma(k_2^*, k_1^*), \qquad j = 1, \ldots, S,$$

which does not depend on j. This means that if we choose a different k^*, the resulting $g_{k^*}(j)$'s differ by only the same constant for all $j \in \mathcal{S}$. With this in mind, we omit the subscript k^* and rewrite (2.9) as

$$\gamma(i,j) = g(j) - g(i), \qquad i,j = 1,\ldots,S. \tag{2.10}$$

$g(i)$ is called the *performance potential* (or simply the *potential*) of state i, and $g_{k^*}(i)$ denotes a particular version of the potential. (The word "potentials" have been used in the literature in similar contents, e.g., [154, 168].) Just as in physics, different versions of the potentials may differ by a constant. Let $g = (g(1),\ldots,g(S))^T$. Then, (2.10) becomes

$$\mathit{\Gamma} = eg^T - ge^T. \tag{2.11}$$

If g is a potential (vector), so is $g + ce$ for any real constant c. For simplicity, we use the same notation g for different versions of the potentials and keep in mind that potential g in different expressions may differ by a constant. A physical interpretation of the performance potentials compared with the potential energy is illustrated in Figure 2.7.

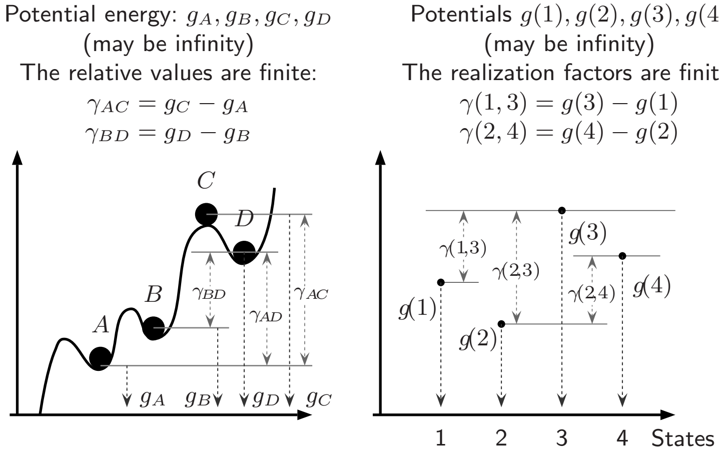

Potential energy: g_A, g_B, g_C, g_D
(may be infinity)
The relative values are finite:
$\gamma_{AC} = g_C - g_A$
$\gamma_{BD} = g_D - g_B$
C
D
B
A
γ_{BD}
γ_{AD}
γ_{AC}
g_A
g_B
g_D
g_C
Potentials $g(1), g(2), g(3), g(4)$
(may be infinity)
The realization factors are finite:
$\gamma(1,3) = g(3) - g(1)$
$\gamma(2,4) = g(4) - g(2)$
$\gamma(1,3)$
$\gamma(2,3)$
$g(3)$
$\gamma(2,4)$
$g(4)$
$g(1)$
$g(2)$
1
2
3
4
States

Fig. 2.7. Physical Interpretation of Potential Energy and Performance Potentials

Substituting (2.11) into (2.7), we obtain

$$e[(I-P)g - f]^T = [(I-P)g - f]e^T,$$

i.e., $e[(I-P)g-f]^T$ is a symmetric matrix. Thus, we must have $(I-P)g-f = ce$, where c is a constant. Left-multiplying both sides of this equation by π and using $\pi = \pi P$, we get $c = -\pi f = -\eta$. Finally, we have

$$(I - P)g + \eta e = f. \tag{2.12}$$

This is called the *Poisson equation.* Its solution is unique only up to an additive constant; i.e., if g is a solution to (2.12), then for any constant c, $g + ce$ is also a solution. To write (2.12) for each component, we have

$$g(i) = f(i) - \eta + \sum_{j \in \mathcal{S}} p(j|i)g(j).$$

This equation has a clear interpretation: The long-term contribution of state i to the average performance, $g(i)$, equals its one-step contribution at the current time, $f(i) - \eta$, plus the expected long-term "potential" contribution of the next state. Equation (2.10) shows that the effect of a perturbation from state i to j (the perturbation realization factor $\gamma(i, j)$) equals the difference in the long-term contributions of these two states.

One of the ways to specify a solution to (2.12) is to normalize it by setting $\pi g = \eta$. In this case, (2.12) takes the form

$$(I - P + e\pi)g = f.$$

It is shown in Appendix B.2 that the eigenvalues of $(I - P + e\pi)$ are $\{1, 1 - \lambda_2, \ldots, 1 - \lambda_S\}$, where λ_i with $|\lambda_i| < 1$, $i = 2, \ldots, S$, are the eigenvalues of the transition probability matrix P [20]. Therefore, $(I - P + e\pi)$ is invertible and the eigenvalues of $(I - P + e\pi)^{-1}$ are $\{1, \frac{1}{1-\lambda_2}, \ldots, \frac{1}{1-\lambda_S}\}$, $|\lambda_i| < 1$, $i = 2, \ldots, S$. Therefore, we have

$$g = (I - P + e\pi)^{-1} f. \tag{2.13}$$

Sample-Path-Based Formulas

The matrix $(I - P + e\pi)^{-1}$ is called the *fundamental matrix* [202]. Because the eigenvalues of $P - e\pi$, $0, \lambda_2, \ldots, \lambda_S$, lie in the unit circle (see Appendix B.2), we can expand the fundamental matrix into a Taylor series:

$$(I - P + e\pi)^{-1} = \sum_{k=0}^{\infty} (P - e\pi)^k = I + \sum_{k=1}^{\infty} (P^k - e\pi). \tag{2.14}$$

Thus, from (2.13), we have

$$g = f + \sum_{k=1}^{\infty} (P^k - e\pi) f.$$

Note that from (A.3), the (i,j)th entry of P^k is $p^{(k)}(j|i) = \mathcal{P}(X_k = j|X_0 = i)$. Then, from (2.14), we have

$$g(i) = \lim_{L\to\infty}\left\{E\left[\sum_{l=0}^{L-1} f(X_l)\Big|X_0 = i\right] - (L-1)\eta\right\}.$$

We may get a more convenient version of the potentials by adding a constant $-\eta$ to every component of g. Thus, we have another version of g:

$$g = [(I - P + e\pi)^{-1} - e\pi]f, \tag{2.15}$$

for which $\pi g = 0$, and

$$\boxed{g(i) = \lim_{L\to\infty} E\left\{\sum_{l=0}^{L-1}[f(X_l) - \eta]\Big|X_0 = i\right\}. \qquad (2.16)}$$

The Poisson equation (2.12) can be easily derived from (2.16); see Problem 2.5.

From (2.16), we can derive another sample-path-based formula for $\gamma(i,j)$. On a sample path of $\boldsymbol{X}$ starting with $X_0 = j$, define $L(i|j)$ to be the first passage time of $\boldsymbol{X}$ reaching state i; i.e., $L(i|j) = \min\{l : l \geq 0, X_l = i|X_0 = j\}$. Then, we have

$$\boxed{\gamma(i,j) = E\left\{\sum_{l=0}^{L(i|j)-1}[f(X_l) - \eta]\Big|X_0 = j\right\}. \qquad (2.17)}$$

This can be intuitively explained as follows: from (2.16),

$$\begin{aligned}
&\gamma(i,j) = g(j) - g(i)\\
&= \lim_{L\to\infty}\left\{E\left\{\sum_{l=0}^{L}[f(X_l) - \eta]\Big|X_0 = j\right\} - E\left\{\sum_{l=0}^{L}\left[f(\widetilde{X}_l) - \eta\right]\Big|\widetilde{X}_0 = i\right\}\right\}\\
&= \lim_{L\to\infty}\left\{E\left\{\left[\sum_{l=0}^{L(i|j)-1}[f(X_l) - \eta] + \sum_{L(i|j)}^{L}[f(X_l) - \eta]\right]\Big|X_0 = j\right\}\right.\\
&\qquad \left. - E\left\{\left[\sum_{l=0}^{L-L(i|j)}\left[f(\widetilde{X}_l) - \eta\right] + \sum_{L-L(i|j)+1}^{L}\left[f(\widetilde{X}_l) - \eta\right]\right]\Big|\widetilde{X}_0 = i\right\}\right\},
\end{aligned} \tag{2.18}$$

where $\{X_l, l \geq 0\}$ and $\{\widetilde{X}_l, l \geq 0\}$ are two independent Markov chains with the same transition probability matrix P. Because of the strong Markov property and $X_{L(i|j)} = i$, the second term equals the third term as long as $L > L(i|j)$. In addition, since $\lim_{l\to\infty} E[f(\widetilde{X}_l)] = \eta$, the last term goes to zero as $L \to \infty$. Thus, (2.18) leads to (2.17). The idea is explained in Figure 2.8: In region (II), both $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}$ are statistically identical because $\widetilde{X}_0 = X_{L(i|j)} = i$; the mean of $f(X_l) - \eta$ on $\widetilde{\boldsymbol{X}}$ in region (III) goes to zero as $L \to \infty$. Thus, the only term left in the difference $g(j) - g(i)$ is the summation on $\boldsymbol{X}$ in region (I). For a detailed proof, see [62].

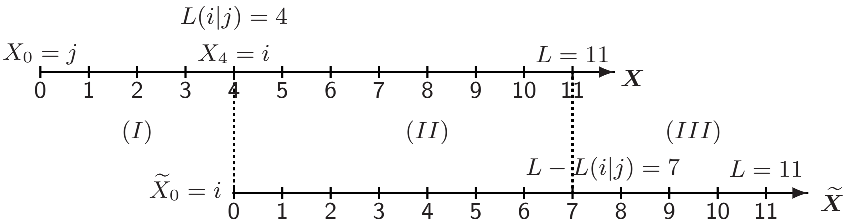


Fig. 2.8. An Explanation for (2.17)

In summary, the perturbation realization factor $\gamma(i, j)$, $i, j \in \mathcal{S}$, measures the "infinitesimal" effect of a perturbation from state i to j on the average reward η (more precisely, it measures the effect on F_L for $L >> 1$). From the physical meaning, the performance potential $g(i)$, $i = 1, \ldots, S$, measures the long-term "potential" contribution of state i to η. Similar to the potential energy in physics, only the differences in the different $g(i)$'s are important for performance sensitivities.

Finally, the reward function can be defined as $f(i, j)$, $i, j \in \mathcal{S}$; i.e, the system gains a reward $f(i, j)$ when it is in state $X_l = i$ and moves to state $X_{l+1} = j$, $l = 0, 1, \ldots$. The average reward is defined as

$$\eta = \lim_{L\to\infty} \frac{1}{L} \sum_{l=0}^{L-1} f(X_l, X_{l+1}).$$

In this case, if we use the average

$$f(i) := \sum_{j=1}^{S} p(j|i) f(i, j)$$

as the reward function, all the results developed in this and the remaining sections for PA of Markov chains hold.

2.1.3 Performance Derivative Formulas

To derive the performance derivative $\frac{d\eta_\delta}{d\delta}$ at policy (P, f) along any direction ΔP, we consider a sample path $\boldsymbol{X}$ with a transition probability matrix P

consisting of L, $L >> 1$, transitions. Among these transitions, on average, there are $L\pi(k)$ transitions at which the system is in state k. Each time when $\boldsymbol{X}$ visits state i after visiting state k, because of the change from P to $P_\delta = P + \delta \Delta P$, the perturbed path $\boldsymbol{X}_\delta$ may have a jump, denoted as from state i to state j (i.e., after visiting k, $\boldsymbol{X}$ moves to i and $\boldsymbol{X}_\delta$ moves to j), as shown in Figure 2.5. For convenience, we allow $i = j$ as a special case. A "real jump" (with $i \neq j$) happens rarely. Denote the probability of a jump from i to j after visiting state k as $p(i,j|k)$. We have

$$p(i,j|k) = p(i|k)p_\delta(k,j|k,i),$$

where $p_\delta(k,j|k,i)$ denotes the conditional probability that $\boldsymbol{X}_\delta$ moves from state k to state j given that $\boldsymbol{X}$ moves from state k to i. By definition, we have $\sum_{j=1}^{S} p_\delta(k,j|k,i) = 1$. Therefore,

$$\sum_{j=1}^{S} p(i,j|k) = p(i|k). \tag{2.19}$$

Similarly,

$$\sum_{i=1}^{S} p(i,j|k) = p_\delta(j|k), \tag{2.20}$$

and $\sum_{i,j=1}^{S} p(i,j|k) = 1$. On average, in the L transitions on the sample path, there are $L\pi(k)p(i,j|k)$ jumps from i to j following the visit to state k. As discussed in Section 2.1.2, each such jump has on average an effect of $\gamma(i,j)$ on F_L.

A real jump happens extremely rarely as $\delta \to 0$. As discussed in Section 2.1.2, the probability that the Markov chain jumps at $l = 3$ and that there is another jump of $\boldsymbol{X}_\delta$ from $l = 4$ to L_{ij}^* (or equivalently, $\boldsymbol{X}_\delta$ would move differently if it followed P from $l = 3$ to L_{ij}^*) is on the order of δ^2; the effect of such a situation can be ignored for performance derivatives. Therefore, we may assume that, from $l = 3$ to L_{ij}^* and in other periods after each jump before merging, $\boldsymbol{X}_\delta$, generated according to P_δ, is the same as following P. Thus, on average, the total effect on F_L due to the change in P to $P_\delta = P + \delta \Delta P$ is

$$\begin{aligned}
&E(F_{\delta,L} - F_L) \\
\approx &\sum_{k=1}^{S} \left[\sum_{i,j=1}^{S} L\pi(k)p(i,j|k)\gamma(i,j) \right] \\
= &\sum_{k=1}^{S} \left\{ \sum_{i,j=1}^{S} L\pi(k)p(i,j|k)[g(j) - g(i)] \right\} \\
= &\sum_{k=1}^{S} L\pi(k) \left\{ \sum_{j=1}^{S} \left[g(j) \sum_{i=1}^{S} p(i,j|k) \right] \right.
\end{aligned}$$

$$-\sum_{i=1}^{S}\left[g(i)\sum_{j=1}^{S}p(i,j|k)\right]\Bigg\}. \tag{2.21}$$

From (2.19) and (2.20), (2.21) becomes

$$\begin{aligned} E(F_{\delta,L}-F_L) &\approx \sum_{k=1}^{S} L\pi(k)\left\{\left[\sum_{j=1}^{S}p_\delta(j|k)g(j)\right]-\left[\sum_{i=1}^{S}p(i|k)g(i)\right]\right\} \\ &= \sum_{k=1}^{S} L\pi(k)\left\{\sum_{j=1}^{S}[p_\delta(j|k)-p(j|k)]g(j)\right\} \\ &= L\pi(P_\delta - P)g = L\pi(\Delta P)\delta g. \end{aligned}$$

Thus,

$$\eta_\delta - \eta = \lim_{L\to\infty}\frac{1}{L}E(F_{\delta,L}-F_L) \approx \pi(\Delta P)\delta g. \tag{2.22}$$

Finally, letting $\delta \to 0$, we obtain the performance derivative formula

$$\boxed{\left.\frac{d\eta_\delta}{d\delta}\right|_{\delta=0} = \pi(\Delta P)g. \qquad (2.23)}$$

Strictly speaking, the approximation in (2.21) is not accurate (the difference of both sides is on the order of $o(L)$, which may not be small for a large L). It is accurate only after both sides of (2.21) are divided by L, resulting in (2.22). Nevertheless, (2.21) provides a good intuition.

From (2.11), we have $\pi\Gamma = g^T - (\pi g)e^T$. Thus, from (2.23), we get

$$\left.\frac{d\eta_\delta}{d\delta}\right|_{\delta=0} = \pi(\Delta P)\Gamma^T\pi^T. \tag{2.24}$$

Note that g, Γ, and π can be estimated on a single sample path of a Markov chain with transition matrix P; thus, given any ΔP, the performance derivative along the direction ΔP can be obtained by (2.23) or (2.24) using the sample path-based estimates of π and g or Γ. Algorithms can be developed for estimating the performance derivative based on a single sample path using (2.23) without estimating each component of g; see Chapter 3.

Finally, (2.23) can be easily derived by using the Poisson equation (2.12). Let P' be the transition probability matrix of another irreducible Markov chain defined on the same state space $\mathcal{S}$, and let η' and π' be its corresponding performance measure and steady-state probability, respectively. Multiplying both sides of (2.12) on the left by π' and using $\pi' e = 1$ and $\pi' = \pi' P'$, we obtain the performance difference formula:

$$\eta' - \eta = \pi'(\Delta P)g. \tag{2.25}$$

Taking P' as $P_\delta = P + \delta(\Delta P)$ and η' as η_δ in (2.25), we get

$$\eta_\delta - \eta = \pi_\delta \delta(\Delta P)g.$$

Letting $\delta \to 0$, we obtain (2.23) (it is easy to see $\lim_{\delta\to 0} \pi_\delta = \pi$). Thus, the performance derivative formula (2.23) follows directly from the Poisson equation (2.12). However, our PA-based reasoning intuitively explains the physical meaning of the realization factors and the potentials. It clearly illustrates the nature of the performance derivatives: They can be constructed by using the potentials as building blocks. More importantly, this PA-based construction approach can be used in constructing performance derivative formulas for other non-standard problems in which the special features of the system can be utilized. New optimization schemes can be developed for such special systems. We discuss these problems in Chapters 8 and 9.

So far, we have assumed that f does not change. Suppose that the reward function associated with P' is f' and, in addition to the change of P to P_δ, f also changes to $f_\delta = f + \delta\Delta f$, $\Delta f = f' - f$. Then, it is easy to obtain the performance derivative formula

$$\left.\frac{d\eta_\delta}{d\delta}\right|_{\delta=0} = \pi[(\Delta P)g + \Delta f]. \tag{2.26}$$

The performance difference formula (2.25) becomes

$$\eta' - \eta = \pi'[(\Delta P)g + \Delta f]. \tag{2.27}$$

The difference between (2.23) (or (2.26)) and (2.25) (or (2.27)) is that π in (2.23) is replaced by π' in (2.25).

With realization factors, we have

$$\left.\frac{d\eta_\delta}{d\delta}\right|_{\delta=0} = \pi\left[(\Delta P)\Gamma^T\pi^T + \Delta f\right] \tag{2.28}$$

and

$$\eta' - \eta = \pi'\left[(\Delta P)\Gamma^T\pi^T + \Delta f\right]. \tag{2.29}$$

Finally, sometimes it may be useful to specifically denote the two policies in performance sensitivity analysis as (P^h, f^h) and (P^d, f^d) (instead of (P', f') and (P, f)). With these notations, (2.26) becomes

$$\left.\frac{d\eta_\delta}{d\delta}\right|_{\delta=0} = \pi^d[(\Delta P)g^d + \Delta f]$$
$$= \pi^d[(P^h - P^d)g^d + (f^h - f^d)],$$

where η_δ is the performance of (P_δ, f_δ), with $P_\delta = P^d + \delta\Delta P$, $\Delta P = P^h - P^d$, $f_\delta = f^d + \delta\Delta f$, and $\Delta f = f^h - f^d$.

2.1.4 Gradients with Discounted Reward Criteria

In this subsection, we show that the idea of performance potentials and the performance derivative formula can be extended to Markov chains with discounted reward criteria.

Consider an ergodic Markov chain $\boldsymbol{X} = \{X_l, l \geq 0\}$ with transition probability matrix P and reward function f. Let β, $0 < \beta \leq 1$, be a discount factor. For $0 < \beta < 1$, we define the discounted reward as a column vector $\eta_\beta = (\eta_\beta(1), \ldots, \eta_\beta(S))^T$ with

$$\eta_\beta(i) := (1-\beta)E\left[\sum_{l=0}^{\infty} \beta^l f(X_l) \Big| X_0 = i\right]. \qquad (2.30)$$

The factor $(1-\beta)$ in (2.30) is used to obtain the continuity of η_β at $\beta = 1$. We show that the long-run average reward discussed in the last subsection can be viewed as a special case when $\beta \to 1$ and therefore we denote $\eta_1 := \eta e$. Also, the weighting factors in (2.30) are normalized: $\sum_{l=0}^{\infty}(1-\beta)\beta^l = 1$. In a matrix form, Equation (2.30) is

$$\eta_\beta = (1-\beta)\sum_{l=0}^{\infty} \beta^l P^l f = (1-\beta)(I - \beta P)^{-1} f, \qquad 0 < \beta < 1. \qquad (2.31)$$

The second equality in (2.31) holds because for $0 < \beta < 1$, all the eigenvalues of βP are within the unit circle [20]. From (2.39) given below, we know that $\lim_{\beta\uparrow 1} \eta_\beta$ exists and we have

$$\eta_1 := \lim_{\beta\uparrow 1} \eta_\beta = \eta e, \qquad (2.32)$$

with $\eta = \pi f$ being the average reward.

$\boldsymbol{\beta}$-Potentials

The *discounted Poisson equation* is defined as

$$\boxed{(I - \beta P + \beta e\pi)g_\beta = f, \qquad 0 < \beta \leq 1. \qquad (2.33)}$$

g_β is called the *β-potential*. When $\beta = 1$, it is the standard Poisson equation (2.12). Thus, the 1−potential is simply the potential (2.13) and is denoted as $g := g_1$. From (2.33), we have

$$\begin{aligned} g_\beta &= (I - \beta P + \beta e\pi)^{-1} f \\ &= \left[\sum_{l=0}^{\infty} \beta^l (P - e\pi)^l \right] f \\ &= \left\{ I + \left[\sum_{l=1}^{\infty} \beta^l (P^l - e\pi) \right] \right\} f, \qquad 0 < \beta \le 1. \end{aligned} \tag{2.34}$$

The above expansion holds because all the eigenvalues of $P - e\pi$ are in the unit circle. In particular, by setting $\beta = 1$ we obtain (2.14).

It is easy to verify the following equations:

$$\pi(I - \beta P + \beta e\pi)^{-1} = \pi, \tag{2.35}$$

$$(I - \beta P + \beta e\pi)^{-1} e = e,$$

$$(I - \beta P)^{-1} e = \frac{1}{1-\beta} e, \tag{2.36}$$

and

$$(I - \beta P)^{-1} = (I - \beta P + \beta e\pi)^{-1} + \frac{\beta}{1-\beta} e\pi. \tag{2.37}$$

Equation (2.37) is obtained by using (2.36), (2.35), and the following equation

$$(I - \beta P)^{-1}(I - \beta P + \beta e\pi) = I + (I - \beta P)^{-1} \beta e\pi.$$

In addition, we have

$$\lim_{\beta \uparrow 1} g_\beta = g_1,$$

$$\pi g_\beta = \pi f. \tag{2.38}$$

From (2.37), we obtain

$$\lim_{\beta \uparrow 1} (1 - \beta)(I - \beta P)^{-1} = e\pi. \tag{2.39}$$

Performance Sensitivities

Suppose that the transition matrix P and the reward function f change to P' and f', respectively, with P' being another irreducible and aperiodic transition matrix. From (2.31), we have

$$\begin{aligned} {\eta'}_\beta - \eta_\beta &= (1-\beta)(f' - f) + \beta(P' {\eta'}_\beta - P\eta_\beta) \\ &= (1-\beta)(f' - f) + \beta(P' - P)\eta_\beta + \beta P'({\eta'}_\beta - \eta_\beta). \end{aligned}$$

This leads to

$$\eta'_{\beta} - \eta_\beta = (1-\beta)(I-\beta P')^{-1}(f'-f) + \beta(I-\beta P')^{-1}(P'-P)\eta_\beta. \tag{2.40}$$

From (2.31) and (2.37), we obtain

$$\eta_\beta = (1-\beta)g_\beta + \beta\eta e. \tag{2.41}$$

Substituting this into the right-hand side of (2.40) and noting that $(P'-P)$ $e=0$, we obtain the *performance difference formula for the discounted reward criterion*:

$$\boxed{\eta'_{\beta} - \eta_\beta = (1-\beta)(I-\beta P')^{-1}[(\beta P' g_\beta + f') - (\beta P g_\beta + f)], \qquad 0<\beta<1. \quad (2.42)}$$

Finally, as a special case, letting $\beta \to 1$ in (2.42) and using (2.39), we obtain the performance difference formula for the long-run average reward (2.27):

$$\eta' - \eta = \pi'[(P'g + f') - (Pg + f)].$$

Now, suppose that P changes to $P_\delta = P + \delta \Delta P$, $\Delta P = P' - P$, and f changes to $f_\delta = f + \delta \Delta f$, $\Delta f = f' - f$, $0 < \delta < 1$. Taking P_δ as the P' in (2.42), we have

$$\eta_{\beta,\delta} - \eta_\beta = (1-\beta)(I-\beta P_\delta)^{-1}[(\beta P_\delta g_\beta + f_\delta) - (\beta P g_\beta + f)], \qquad 0<\beta<1. \tag{2.43}$$

Letting $\delta \downarrow 0$, we obtain the *performance derivative formula for the discounted reward criterion:*

$$\boxed{\left.\frac{d\eta_{\beta,\delta}}{d\delta}\right|_{\delta=0} = (1-\beta)(I-\beta P)^{-1}(\beta \Delta P g_\beta + \Delta f), \qquad 0<\beta<1. \quad (2.44)}$$

When $\beta \uparrow 1$, this equation reduces to (2.26).

From (2.41) and (2.44), we have

$$\boxed{\left.\frac{d\eta_{\beta,\delta}}{d\delta}\right|_{\delta=0} = (I-\beta P)^{-1}[\beta \Delta P \eta_\beta + (1-\beta)\Delta f], \qquad 0<\beta<1.}$$

Similarly, from (2.41) and (2.43), we have

$$\boxed{\eta_{\beta,\delta} - \eta_\beta = (I - \beta P_\delta)^{-1}[\beta \Delta P \eta_\beta + (1-\beta)\Delta f], \qquad 0 < \beta < 1.}$$

All the applications of the performance potentials g_β in optimization depend only on the differences of the components of g_β. In other words, we can replace g_β with $g_\beta + ce$, where c is any constant. These are different versions of the β-potential, and for simplicity, we will use the same notation g_β to denote them. In particular, the performance difference and derivative formulas (2.42) and (2.44) hold when g_β is replaced by $g_\beta + ce$. Therefore, we may add a constant vector $-\eta e$ to (2.34) and obtain a sample-path-based expression for the β-potential (cf., (2.16)) as follows:

$$\begin{aligned} g_\beta(i) &= \lim_{L\to\infty} E\left\{ \sum_{l=0}^{L-1} \beta^l [f(X_l) - \eta] \Big| X_0 = i \right\} \\ &= E\left\{ \sum_{l=0}^{\infty} \beta^l [f(X_l) - \eta] \Big| X_0 = i \right\}, \qquad 0 < \beta \leq 1, \end{aligned} \tag{2.45}$$

in which we have exchanged the order of $\lim_{L\to\infty}$ and "E". Of course, for $0 < \beta < 1$, we can also discard the constant term $(\sum_{l=0}^{\infty} \beta^l)\eta$ and obtain

$$\boxed{g_\beta(i) = E\left\{ \left[\sum_{l=0}^{\infty} \beta^l f(X_l) \right] \Big| X_0 = i \right\}. \qquad (2.46)}$$

In modern Markov theory, (2.46) is called the β-potential of reward function f with $0 < \beta < 1$. (In [87], it is called the α-potential, since α is used as the discount factor there; in this book, we reserve α to denote actions in MDPs.) Therefore, from (2.16), (2.45), and (2.46), the potential for the long-run average reward, $g(i)$, is a natural extension of the β-potential from $0 < \beta < 1$ to $\beta = 1$; a constant η is subtracted from each term in (2.46) to keep the sum finite when extended to $\beta = 1$. This justifies again our terminology of "potential" for g in the long-run average reward case.

It is clear that the β-potential (2.46) is almost the same as the discounted reward (2.30). This explains why the concept of the discounted performance potential is not introduced in many previous works on the optimization of discounted rewards. Nevertheless, this concept puts the approach to the discounted-reward optimization problem in the same framework as the approach to the average-reward problem. This is also true for the policy iteration approach in MDPs, see Chapter 4.

Similar to the average-reward case, we define the (discounted) *PRF matrix* as

$$\Gamma_\beta = e g_\beta^T - g_\beta e^T, \qquad 0 < \beta \le 1.$$

From this equation and (2.38), we have

$$\Gamma_\beta^T \pi^T = g_\beta - \eta e.$$

The (discounted) PRF matrix satisfies the *(discounted) PRF equations:*

$$-\Gamma_\beta + \beta P \Gamma_\beta P^T = -F, \tag{2.47}$$

where $F = e f^T - f e^T$.

This can be easily verified:

$$\begin{aligned} -\Gamma_\beta + \beta P \Gamma_\beta P^T &= -(e g_\beta^T - g_\beta e^T) + \beta P (e g_\beta^T - g_\beta e^T) P^T \\ &= -(e g_\beta^T - g_\beta e^T) + \beta [e (P g_\beta)^T - P g_\beta e^T] \\ &= -[e(g_\beta - \beta P g_\beta + \beta e \pi g_\beta)^T - (g_\beta - \beta P g_\beta + \beta e \pi g_\beta) e^T] \\ &= -F. \end{aligned}$$

Equation (2.47) reduces to the standard PRF equation (2.7) when $\beta = 1$.

With the PRF matrix, (2.42) and (2.44) become

$$\eta'_\beta - \eta_\beta = (1-\beta)(I - \beta P')^{-1} [\beta (\Delta P) \Gamma_\beta^T \pi^T + \Delta f], \qquad 0 < \beta < 1,$$

and

$$\frac{d\eta_{\beta,\delta}}{d\delta} = (1-\beta)(I - \beta P)^{-1} [\beta (\Delta P) \Gamma_\beta^T \pi^T + \Delta f], \qquad 0 < \beta < 1.$$

Again, when $\beta \uparrow 1$, these two sensitivity formulas reduce to the average-reward case (2.28) and (2.29).

Figure 2.9 summarizes the results for both the discounted- and average-reward performance sensitivity analysis with a unified view; all the results of the average-reward case can be obtained by setting $\beta \uparrow 1$ from those of the discounted-reward case.

Intuitions

Finally, we offer an intuitive explanation for the discounted reward derivative formula (2.44). For simplicity, we assume that $\Delta f = 0$. Because the discounted reward $\eta_\beta(i)$ depends on the initial state $i \in \mathcal{S}$, we have to consider the transient probabilities on the sample path. Consider a sample path $\boldsymbol{X} = \{X_0, X_1, \ldots\}$ starting from $X_0 = i \in \mathcal{S}$. The conditional probability of $X_l = k$, $l = 1, 2, \ldots$, given that $X_0 = i$ is $\mathcal{P}(X_l = k | X_0 = i) = p^{(l)}(k|i)$ (cf. (A.3)). Let $p_l(u, v|k)$ be the probability that, given $X_l = k$, the system has a jump from

β	Discounted (0,1) $\longrightarrow$	Average 1
Performance	$\eta_\beta(i)=(1-\beta)\mathrm{E}\Big[\sum_{l=0}^{\infty}\beta^l f(X_l)\Big\vert X_0=i\Big]$ $\eta_\beta=(\eta_\beta(1),\cdots,\eta_\beta(S))^{\mathrm{T}}$ $\longrightarrow$	$\eta=\lim_{N\to\infty}\frac{1}{N}\mathrm{E}\Big[\sum_{l=0}^{N-1}f(X_l)\Big\vert X_0=i\Big]$ ηe
Potentials	$g_\beta(i)=\mathrm{E}\sum_{l=0}^{\infty}\Big\{\beta^l[f(X_l)-\eta]\Big\vert X_0=i\Big\}$ or $g_\beta(i)=\mathrm{E}\Big[\sum_{l=0}^{\infty}\beta^l f(X_l)\Big\vert X_0=i\Big]$ $\longrightarrow$	$g(i)=\mathrm{E}\sum_{l=0}^{\infty}\Big\{[f(X_l)-\eta]\Big\vert X_0=i\Big\}$
Poisson Eq.	$(I-\beta P+\beta e\pi)g_\beta=f$ $\longrightarrow$	$(I-P+e\pi)g=f$
Realization factors	$\Gamma_\beta=eg_\beta^{\mathrm{T}}-g_\beta e^{\mathrm{T}}$ $-\Gamma_\beta+\beta P\Gamma_\beta P^{\mathrm{T}}=-F$ $\longrightarrow$	$\Gamma=eg^{\mathrm{T}}-ge^{\mathrm{T}}$ $-\Gamma+P\Gamma P^{\mathrm{T}}=-F$
Performance Difference	$(1-\beta)(I-\beta P)^{-1}$ $\longrightarrow$ $\eta'_\beta-\eta_\beta=(1-\beta)(I-\beta P)^{-1}$ $[(\beta P'g_\beta+f')-(\beta Pg_\beta+f)]$ $\longrightarrow$	$e\pi$ $\eta'-\eta=\pi'[(P'g+f')-(Pg+f)]$
Performance Derivative	$\frac{\mathrm{d}\eta_\beta}{\mathrm{d}\delta}=(1-\beta)(I-\beta P)^{-1}(\beta\Delta Pg_\beta+\Delta f)$	$\frac{\mathrm{d}\eta}{\mathrm{d}\delta}=\pi(\Delta Pg+\Delta f)$

Fig. 2.9. A Comparison of Discounted- and Average-Reward Problems

state u to v at time $l+1$ (i.e., the original system with transition probability matrix P moves from $X_l=k$ to $X_{l+1}=u$, but the perturbed system with $P_\delta=P+\delta(\Delta P)$ moves from $X_l=k$ to $X_{l+1}=v$). The effect of such a jump, measured starting from $l+1$, is $\gamma_\beta(u,v)=g_\beta(v)-g_\beta(u)$. Since the jump happens at time $l+1$, its effect on the discounted reward $\eta_\beta(i)$ in (2.30) is $\beta^{l+1}\gamma_\beta(u,v)$. Therefore, from the physical meaning, we can decompose $\Delta\eta_\beta(i)$ into

$$\begin{aligned}\Delta\eta_\beta(i) &= (1-\beta)\left[\sum_{l=0}^{\infty}\sum_{k=1}^{S}\sum_{u=1}^{S}\sum_{v=1}^{S}\beta^{l+1}\mathcal{P}(X_l=k|X_0=i)p_l(u,v|k)\gamma_\beta(u,v)\right]\\ &= (1-\beta)\left\{\sum_{l=0}^{\infty}\sum_{k=1}^{S}\beta^{l+1}p^{(l)}(k|i)\sum_{u=1}^{S}\sum_{v=1}^{S}\{p_l(u,v|k)[g_\beta(v)-g_\beta(u)]\}\right\}.\end{aligned}$$

Similar to (2.19) and (2.20), we have

$$\sum_{v=1}^{S} p_l(u,v|k) = p(u|k) \quad \text{and} \quad \sum_{u=1}^{S} p_l(u,v|k) = p_\delta(u|k).$$

Thus,

$$\Delta\eta_\beta(i) = (1-\beta)\left\{\sum_{l=0}^{\infty}\sum_{k=1}^{S}\beta^{l+1}p^{(l)}(k|i)\left\{\sum_{j=1}^{S}[p_\delta(j|k)-p(j|k)]g_\beta(j)\right\}\right\}.$$

In a matrix form, this is

$$\begin{aligned}\Delta\eta_\beta &= (1-\beta)\left\{\sum_{l=0}^{\infty}\beta^{l+1}P^l[(\Delta P)\delta g_\beta]\right\}\\ &= (1-\beta)(I-\beta P)^{-1}[\beta(\Delta P)\delta g_\beta],\end{aligned}$$

which directly leads to (2.44).

2.1.5 Higher-Order Derivatives and the MacLaurin Series

In this section, we continue our study by exploring the system's behavior in the neighborhood of a given policy P in the policy space.

Higher-Order Derivatives

We assume that P changes to $P_\delta = P + \delta(\Delta P)$, $\Delta P = P' - P$, and we let $f_\delta \equiv f$, for simplicity. Denote $B = P - I$, which can be viewed as an infinitesimal generator of a Markov process with unit transition rates and a transition probability matrix P for its embedded chain (see Appendix A.2). To study the higher-order derivatives with respect to δ, it is convenient to use short-hand notation defined as

$$\begin{aligned}B^{\#} &= -(-B+e\pi)^{-1}+e\pi\\ &= -[(I-P+e\pi)^{-1}-e\pi].\end{aligned} \tag{2.48}$$

$B^{\#}$ is called the *group inverse* of B [202], which satisfies

$$BB^{\#} = B^{\#}B = I - e\pi, \tag{2.49}$$

and

$$B^{\#}e = 0, \qquad \pi B^{\#} = 0.$$

The term "group" comes from the following fact. For any probability distribution π on state space $\mathcal{S}$, define a set of $S \times S$ matrices

$$\mathcal{B} := \{B :\ \pi B = 0,\ Be = 0\}. \tag{2.50}$$

It is easy to verify that $\mathcal{B}$ is a group (see, e.g., [219] for a definition) with identity element $I - e\pi$ under the operation of matrix multiplication (see Problem 2.11). Equation (2.49) indicates that $B^{\#}$ is indeed the inverse of B in group $\mathcal{B}$.

With the group inverse, the potential in (2.15) becomes

$$g = -B^{\#} f,$$

and the performance derivative formula (2.23) takes the form

$$\left.\frac{d\eta_\delta}{d\delta}\right|_{\delta=0} = \pi(\Delta P)g = \pi[(\Delta P)(-B^{\#})]f.$$

For the irreducible finite Markov chain with transition matrix P_δ, we have

$$\pi_\delta(I - P_\delta) = 0,$$

and $\frac{dP_\delta}{d\delta} = \Delta P$. By taking derivatives on both sides of this equation with respect to δ, we have

$$\frac{d\pi_\delta}{d\delta}(I - P_\delta) = \pi_\delta(\Delta P).$$

Continuously taking derivatives on both sides of the resulting equations, we obtain for any $n \geq 1$,

$$\frac{d^n \pi_\delta}{d\delta^n}(I - P_\delta) = n \frac{d^{n-1}\pi_\delta}{d\delta^{n-1}}(\Delta P).$$

Setting $\delta = 0$ and multiplying both sides of the above equation on the right by $-B^{\#}$ and noting that $BB^{\#} = I - e\pi$ and $\pi e = 1$, we get

$$\left.\frac{d^n \pi_\delta}{d\delta^n}\right|_{\delta=0} = n \left.\frac{d^{n-1}\pi_\delta}{d\delta^{n-1}}\right|_{\delta=0} [(\Delta P)(-B^{\#})].$$

Thus,

$$\left.\frac{d^n \pi_\delta}{d\delta^n}\right|_{\delta=0} = n!\pi[(\Delta P)(-B^{\#})]^n.$$

Finally, for any reward function f, we have

$$\boxed{\begin{aligned}\left.\frac{d^n \eta_\delta}{d\delta^n}\right|_{\delta=0} &= n!\pi[(\Delta P)(-B^{\#})]^n f \\ &= n!\pi[(\Delta P)(I - P + e\pi)^{-1}]^n f, \qquad n \geq 1.\end{aligned}}$$

The MacLaurin Expansion

We note that η_δ is an analytical function of δ. (More precisely, it is a rational function of δ whose denominator and numerator are both polynomials of δ with finite degrees. This can be verified by solving $\pi_\delta(I-P_\delta)=0$ and $\pi_\delta e=1$.) Thus, η_δ has a MacLaurin expansion at $\delta=0$:

$$\begin{aligned}\eta_\delta-\eta &= \sum_{n=1}^{\infty}\frac{1}{n!}\left.\frac{d^n\eta_\delta}{d\delta^n}\right|_{\delta=0}\delta^n\\ &= \pi\left\{\sum_{n=1}^{\infty}[(\varDelta P)(-B^{\#})]^n\delta^n\right\}f, \qquad (2.51)\end{aligned}$$

or equivalently,

$$\boxed{\eta_\delta=\pi\sum_{n=0}^{\infty}\{[(\varDelta P)(-B^{\#})]^n f\delta^n\}. \qquad (2.52)}$$

Denote the spectrum radius of a matrix W as $\rho(W)$ (i.e., the largest absolute value of the eigenvalues of W). Define

$$r=\frac{1}{\rho[(\varDelta P)(-B^{\#})]}=\frac{1}{\rho[(\varDelta P)(I-P+e\pi)^{-1}]}.$$

Then, for $\delta<r$, the eigenvalues of $\delta(\varDelta P)B^{\#}$ are all in the unit circle, and the summation in (2.52) converges. Therefore, for $\delta<r$, we have

$$\begin{aligned}\sum_{n=0}^{\infty}[(\varDelta P)(-B^{\#})\delta]^n &= [I-\delta(\varDelta P)(-B^{\#})]^{-1}\\ &= [I-\delta(\varDelta P)(I-P+e\pi)^{-1}]^{-1}. \qquad (2.53)\end{aligned}$$

Next, if we take $f=e_{\cdot i}$, where $e_{\cdot i}$ is a column vector representing the ith column of the identity matrix I, then the corresponding performance is $\pi_\delta e_{\cdot i}=\pi_\delta(i)$, $i\in\mathcal{S}$. Thus, from (2.52), we have

$$\pi_\delta(i)=\pi\sum_{n=0}^{\infty}\{[(\varDelta P)(-B^{\#})]^n e_{\cdot i}\delta^n\}.$$

In matrix form, we have

$$\pi_\delta=\pi\sum_{n=0}^{\infty}\{[(\varDelta P)(-B^{\#})]^n\delta^n\}, \qquad \delta<r. \qquad (2.54)$$

Thus, from (2.51) we obtain

$$\eta_\delta - \eta = \pi_\delta \delta(\mathit{\Delta} P)(-B^{\#})f.$$

This is consistent with the performance difference formula (2.25). From (2.52), (2.53), and (2.54), we establish a general form:

$$\boxed{\begin{aligned}\eta_\delta &= \pi \sum_{k=0}^{n} [\delta(\mathit{\Delta} P)(-B^{\#})]^k f + \pi \sum_{k=n+1}^{\infty} [\delta(\mathit{\Delta} P)(-B^{\#})]^k f \\ &= \pi \sum_{k=0}^{n} [\delta(\mathit{\Delta} P)(-B^{\#})]^k f + \pi_\delta [\delta(\mathit{\Delta} P)(-B^{\#})]^{n+1} f, \\ &\qquad\qquad \delta < r, \qquad \text{for any } n \geq 0. \qquad (2.55)\end{aligned}}$$

The last term in (2.55), $\pi_\delta[\delta(\mathit{\Delta} P)(-B^{\#})]^{n+1} f$, is the error in taking the first $(n+1)$ terms in the MacLaurin series as an estimate of η_δ. Equations (2.54) and (2.55) hold for $\delta < r$. If $r > 1$, then we can set $\delta = 1$ in (2.54) and (2.55) and obtain the performance value for $P' = P + \mathit{\Delta} P$ as follows

$$\begin{aligned}\eta' &= \pi \sum_{k=0}^{\infty} [(\mathit{\Delta} P)(-B^{\#})]^k f \\ &= \pi \sum_{k=0}^{n} [(\mathit{\Delta} P)(-B^{\#})]^k f + \pi' [(\mathit{\Delta} P)(-B^{\#})]^{n+1} f, \qquad \text{for any } n \geq 0.\end{aligned}$$

The extensions to Markov chains with general state space are in [131, 133].

Numerical Calculations

The saving in computation is significant when we use the MacLaurin series to calculate the performance for many different $(\mathit{\Delta} P)$'s and δ's. There is only one matrix inversion $(I - P + e\pi)^{-1}$ involved. The nth derivative of π_δ at $\delta = 0$, i.e., $n!\pi[(\mathit{\Delta} P)(I - P + e\pi)^{-1}]^n$, can be simply obtained by multiplying the $(n-1)$th derivative, i.e., $(n-1)!\pi[(\mathit{\Delta} P)(I - P + e\pi)^{-1}]^{n-1}$, with the matrix $n(\mathit{\Delta} P)(I - P + e\pi)^{-1}$. For example, π_δ in (2.54) can be calculated simply as follows. First, we set $G_\delta := \delta(\mathit{\Delta} P)(I - P + e\pi)^{-1}$. Then, we

i. solve $\pi = \pi P$ and $\pi e = 1$ to obtain π, calculate G_δ, and set $\pi_\delta := \pi$;
ii. recursively calculate $\pi_\delta := \pi + \pi_\delta G_\delta$, until π_δ reaches a desired precision.

The matrix $(I - P + e\pi)^{-1}$ can be estimated by analyzing a sample path of the Markov chain (see, e.g., Problem 3.18 in Chapter 3). Thus, with a sample-path-based approach, matrix inversion is not even needed. This also implies that, principally, we can obtain the performance of a Markov system with any transition probability matrix P_δ by analyzing a sample path of the Markov system with transition probability matrix P as long as $\delta < r$.

It is not easy to determine the value of r. However, there exist some upper bounds (albeit not tight) for $\rho[\Delta P(-B^{\#})] = \frac{1}{r}$ (or lower bounds for r). From spectrum theory, we have [20]

$$\rho[\Delta P(I - P + e\pi)^{-1}] \le ||\Delta P|| \times ||(I - P + e\pi)^{-1}||,$$

where $||\cdot||$ denotes the norm of a matrix, which is defined as

$$||\Delta P|| = \max_{\text{all } i} \sum_j |\Delta p(j|i)|.$$

Thus, if $||\Delta P||$ is not large, then the spectrum radius of $(\Delta P)(-B^{\#})$ may be small. Let $s^+ = \max_{\text{all } i} \sum_{j:\ \Delta p(i,j)>0} [\Delta p(j|i)]$. If $s^+ \le 0.5$, then $||\Delta P|| \le 1$. Since $\Delta P = P' - P$, at least we have $||\Delta P|| \le 2$.

It is, however, not easy to obtain the norm of the fundamental matrix. As shown in Section 2.1.2, the eigenvalues of $(I - P + e\pi)^{-1}$ are $\{1, \frac{1}{1-\lambda_2}, \ldots, \frac{1}{1-\lambda_S}\}$, with $|\lambda_i| < 1$, $i = 2, \ldots, S$, and 1 being the eigenvalues of P. Thus, we have

$$\rho(I - P + e\pi)^{-1} = \frac{1}{\inf\{1,\ |1-\lambda_i|, i = 2, \ldots, S\}}.$$

However, there is no direct link between $\rho(I - P + e\pi)^{-1}$ and $\rho\Delta P(I - P + e\pi)^{-1}$.

The worst case happens when P has an eigenvalue that is close to 1. If $P \approx I$ (I is a reducible matrix and hence cannot be chosen as P), then $I - P + e\pi \approx e\pi$, which has an eigenvalue 0. The radius of $(I - P + e\pi)^{-1}$ is thus very large. In this case, the radius of $\Delta P(I - P + e\pi)^{-1}$ may be also very large; i.e., r may be very small.

	1	2	3	4	5
1	0.000	0.300	0.200	0.100	0.400
2	0.500	0.000	0.000	0.300	0.200
3	0.200	0.150	0.000	0.150	0.500
4	0.400	0.200	0.150	0.150	0.100
5	0.250	0.250	0.250	0.250	0.000

Table 2.1. The Matrix P

Example 2.1. For illustrative purposes, we study a Markov chain with five states. The state transition matrix P is listed in Table 2.1; the change of P, ΔP, is listed in Table 2.2; and the reward function f is given in Table 2.3. All the values are arbitrarily chosen with some considerations about generality. Note that ΔP represents some dramatic changes in P, e.g., $p(1|2)$ changes

	1	2	3	4	5
1	0.100	-0.300	0.100	0.000	0.100
2	-0.500	0.000	0.500	0.000	0.000
3	-0.200	0.100	0.000	0.100	0.000
4	-0.100	0.100	-0.050	0.000	0.050
5	-0.250	0.250	0.000	0.000	0.000

Table 2.2. The Matrix ΔP

	1	2	3	4	5
f	10	5	1	15	3
π	0.256	0.192	0.136	0.189	0.228

Table 2.3. f and π

	1	2	3	4	5
1	0.803	0.077	0.048	-0.055	0.127
2	0.180	0.858	-0.094	0.079	-0.023
3	-0.048	-0.026	0.900	-0.021	0.196
4	0.112	0.012	0.008	0.949	-0.082
5	0.006	0.039	0.079	0.049	0.827

Table 2.4. The Matrix $(I - P + e\pi)^{-1}$

	1	2	3	4	5
1	0.022	-0.249	0.131	-0.026	0.122
2	-0.425	-0.052	0.426	0.017	0.034
3	-0.131	0.072	-0.018	0.114	-0.036
4	-0.059	0.081	-0.055	0.017	0.017
5	-0.156	0.195	-0.036	0.033	-0.038

Table 2.5. The Matrix $\Delta P(I - P + e\pi)^{-1}$

from 0.5 to 0, and $p(3|2)$ changes from 0 to 0.5. We calculated the matrices $(I - P + e\pi)^{-1}$ and $\Delta P(I - P + e\pi)^{-1}$, which are listed in Tables 2.4 and 2.5. The eigenvalues of $\Delta P(I - P + e\pi)^{-1}$ are 0.3176, -0.3415, 0, -0.0164 +0.0325i, and -0.0164-0.0325i; all of them are inside the unit circle. Thus, $r > 1$ and the MacLaurin series converges within $\delta \leq 1$. ($\delta > 1$ does not make sense, since for $\delta > 1$, $p_\delta(1|2) < 0$). Table 2.6 lists the coefficients of the first to the tenth terms in the MacLaurin series, i.e., $\pi[\Delta P(I - P + e\pi)^{-1}]^n$, for $n = 1, 2, \ldots, 10$. The coefficients of the terms with orders higher than 10 are all numerically zeros. Table 2.7 lists the performance values of the Markov chains with $P_\delta = P + \delta\Delta P$, $\delta = 0.1, 0.2, \ldots, 0.9, 1$, obtained by using the first n terms of the MacLaurin series, $n = 1, 2, \ldots, 10$. All these values converge

	1	2	3	4	5
1st	-0.14050	-0.00387	0.09417	0.02282	0.02738
2nd	-0.01941	0.04907	-0.02399	0.01562	-0.02129
3th	-0.01577	-0.00232	0.01869	-0.00184	0.00123
4th	-0.00189	0.00547	-0.00334	0.00251	-0.00275
5th	-0.00165	-0.00038	0.00210	-0.00029	0.00022
6th	-0.00017	0.00060	-0.00041	0.00028	-0.00030
7th	-0.00017	-0.00006	0.00024	-0.00004	0.00003
8th	-0.00001	0.00007	-0.00005	0.00003	-0.00003
9th	-0.00002	-0.00001	0.00003	0.00000	0.00000
10th	0.00000	0.00001	-0.00001	0.00000	0.00000

Table 2.6. The Coefficients of the MacLaurin Series of η_δ

to the actual average reward of the corresponding Markov chains. Note that after $n = 6$ the values change very little. The average reward of the original Markov chain ($\delta = 0$) is 7.1647. □

n δ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1	7.0741	6.9836	6.8930	6.8024	6.7118	6.6212	6.5307	6.4401	6.3495	6.2589
2	7.0761	6.9915	6.9108	6.8340	6.7612	6.6924	6.6275	6.5666	6.5096	6.4566
3	7.0759	6.9901	6.9061	6.8229	6.7394	6.6547	6.5677	6.4773	6.3825	6.2822
4	7.0759	6.9901	6.9063	6.8237	6.7416	6.6592	6.5760	6.4914	6.4051	6.3166
5	7.0759	6.9901	6.9063	6.8235	6.7410	6.6576	6.5726	6.4849	6.3933	6.2967
6	7.0759	6.9901	6.9063	6.8236	6.7410	6.6578	6.5731	6.4860	6.3955	6.3009
7	7.0759	6.9901	6.9063	6.8236	6.7410	6.6578	6.5729	6.4855	6.3944	6.2986
8	7.0759	6.9901	6.9063	6.8236	6.7410	6.6578	6.5729	6.4856	6.3946	6.2991
9	7.0759	6.9901	6.9063	6.8236	6.7410	6.6578	6.5729	6.4855	6.3945	6.2988
10	7.0759	6.9901	6.9063	6.8236	6.7410	6.6578	6.5729	6.4855	6.3946	6.2989

Table 2.7. The Performance Calculated by the MacLaurin Series

The next example shows that r may be less than 1.

Example 2.2. Consider

$$P = \begin{bmatrix} 0.90 & 0.10 \\ 0.15 & 0.85 \end{bmatrix},$$

$$\Delta P = \begin{bmatrix} -0.8 & 0.8 \\ 0.8 & -0.8 \end{bmatrix},$$

and $f = (1, 5)^T$. Then we have $\pi = (0.6, 0.4)$, and

$$\Delta P (I - P + e\pi)^{-1} = \begin{bmatrix} -3.2 & 3.2 \\ 3.2 & -3.2 \end{bmatrix}.$$

Its eigenvalues are 0 and -6.4. Therefore, the MacLaurin series converges only if $\delta < \frac{1}{|-6.4|} = 0.156$. In fact, as n increases, $\pi[\varDelta P(I - P + e\pi)^{-1}]^n$ goes to infinity very rapidly. Note that the matrix P in this example is close to I.

The curve in Figure 2.10 shows the performance of the system for $\delta \in [0, 1]$. The five points $(*)$ in the figure show the performance of the system calculated by the MacLaurin series corresponding to $\delta = 0.03, 0.06, 0.09, 0.12$, and 0.15. The first four points are the values given by the first 10 terms of the MacLaurin series, and the fifth point is given by 50 terms. At the first three points ($\delta = 0.03, 0.06, 0.09$), the MacLaurin series almost reaches the true value after the first 10 terms (with an error of less than 0.001). The last point is very close to the convergence range ($\delta = 0.15 \approx r = 0.156$), and as shown in the figure, the MacLaurin series does not converge even after 50 terms. In fact, it does converge after 200 terms. □

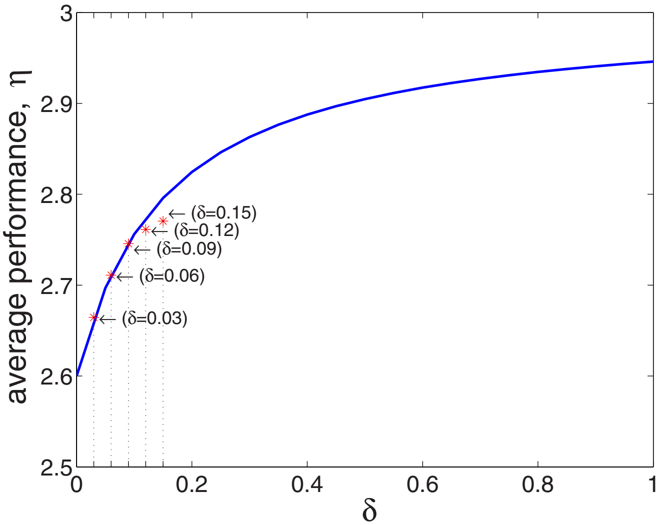


Fig. 2.10. The Performance Compared with MacLaurin Series

Extension to General Function P_θ

Now, let us extend the results to the more general case when the transition probability matrix is a function of θ denoted as P_θ. We assume that the first and all the higher-order derivatives of P_θ with respect to θ exist at $\theta = 0$. Set $P_0 = P$ and $\varDelta P_\theta := P_\theta - P$. Let the reward function $f_\theta \equiv f$ for all θ. Let π_θ be the steady-state probability vector of the Markov chain with transition probability matrix P_θ, and η_θ be its corresponding long-run average reward.

We may use (2.52) to get an expansion of η_θ. For any fixed $\theta > 0$, we simply set $\Delta P = \Delta P_\theta$ and $\delta = 1$ in (2.52). Assume that θ is small enough so that $\rho[\Delta P_\theta(-B^{\#})] < 1$ and therefore expansion (2.52) exists. Then, we have

$$\eta_\theta = \pi \sum_{n=0}^{\infty} \{[\Delta P_\theta(-B^{\#})]^n f\}. \tag{2.56}$$

Equation (2.55) becomes

$$\boxed{\eta_\theta = \pi \sum_{k=0}^{n} [\Delta P_\theta(-B^{\#})]^k f + \pi_\theta [\Delta P_\theta(-B^{\#})]^{n+1} f, \qquad \text{for any } n \geq 0.}$$

Note that this expansion is not a MacLaurin series of η_θ in terms of θ. In fact, ΔP_θ has an expansion

$$\Delta P_\theta = \left.\frac{dP_\theta}{d\theta}\right|_{\delta=0} \theta + \frac{1}{2!} \left.\frac{d^2 P_\theta}{d\theta^2}\right|_{\delta=0} \theta^2 + \cdots,$$

where the derivatives are taken at $\theta = 0$. Substituting it into (2.56), we obtain the MacLaurin series of η_θ:

$$\eta_\theta = \pi \left\{ I + \left[\frac{dP_\theta}{d\theta}(-B^{\#}) \right] \theta + \left\{ \frac{1}{2!} \frac{d^2 P_\theta}{d\theta^2}(-B^{\#}) + \left[\frac{dP_\theta}{d\theta}(-B^{\#}) \right]^2 \right\} \theta^2 + \cdots \right\} f. \tag{2.57}$$

Therefore, we have

$$\frac{d\eta_\theta}{d\theta} = \pi \frac{dP_\theta}{d\theta}(-B^{\#}) f = \pi \frac{dP_\theta}{d\theta} g \tag{2.58}$$

and

$$\frac{d^2 \eta_\theta}{d\theta^2} = \pi \left\{ \frac{d^2 P_\theta}{d\theta^2}(-B^{\#}) + 2! \left[\frac{dP_\theta}{d\theta}(-B^{\#}) \right]^2 \right\} f.$$

Other higher-order derivatives can be obtained in a similar way.

Benes [19] presented an interesting result on the MacLaurin series of the call blocking probability in terms of the input call intensity λ in a telecommunication network. The results presented in this section are more general and concise and can be applied on-line when the system is running. Other related works are in [29], [120], [153], [158], and [267]. In [120], the MacLaurin series of the moments of the response times in a GI/G/1 queue is derived; the results are extended to inventory systems in [29], [153], [158]; and [267] focuses on the expansion of performance measures in queueing systems.

2.2 Performance Sensitivities of Markov Processes

In this section, we extend the aforementioned sensitivity analysis results to (continuous-time) Markov processes. Consider an irreducible and aperiodic (ergodic) Markov process $\boldsymbol{X} = \{X_t, t \geq 0\}$ with a finite state space $\mathcal{S} = \{1, 2, \ldots, S\}$ and an infinitesimal generator $B = [b(i,j)]$, where $b(i,j) \geq 0,\ i \neq j,\ b(i,i) < 0$. Let π be the steady-state probability (row) vector. We have $\pi e = 1$ and

$$Be = 0, \qquad \pi B = 0.$$

We can construct an embedded Markov chain (discrete-time) that has the same steady-state probability as the Markov process $\boldsymbol{X}$. This is called *uniformization* (see Problem A.8). Thus, the sensitivity analysis of a Markov process can be converted to that of a Markov chain, and then the results in Section 2.1 can be translated to Markov processes. In this section, however, we adopt a direct approach, which provides a clear meaning and intuition.

Perturbation Realization

Let f be a *reward function* on $\mathcal{S}$ and also denote a (column) vector $f = (f(1), \ldots, f(S))^T$. The long-run average performance measure of the Markov process is:

$$\eta = \pi f = \lim_{T \to \infty} \frac{1}{T} E \left[\int_0^T f(X_t) dt \right],$$

which exists for ergodic Markov processes, where E denotes the expectation.

To determine the effect of a perturbation (jump) from state i to state j on the performance η, we study two independent sample paths $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}$ with the same infinitesimal generator B, starting from initial states $X_0 = i$ and $\widetilde{X}_0 = j$, respectively. Let E denote the expectation in the probability space spanned by all the sample paths of both $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}$. By the ergodicity of $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}$, they will merge together with probability 1. Define

$$\boxed{T^*_{ij} = \inf \left\{ t : t \geq 0, X_t = \widetilde{X}_t \,\middle|\, X_0 = i, \widetilde{X}_0 = j \right\}.}$$

By the strong Markov property, after T^*_{ij}, the two processes $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}$ will behave similarly probabilistically. T^*_{ij} is just the coupling time of the two independent Markov processes with different initial states. Readers are referred to [203] for a survey of the relevant results about coupling.

Now, we define the *perturbation realization factor (PRF)* as (cf. (2.5))

$$\gamma(i,j) = E\left\{ \int_0^{T_{ij}^*} [f(\widetilde{X}_t) - f(X_t)]dt \,\middle|\, X_0 = i, \widetilde{X}_0 = j \right\}, \qquad i,j \in \mathcal{S}. \tag{2.59}$$

The PRF matrix is $\Gamma := [\gamma(i,j)]$. From the definition, we have

$$\gamma(i,j) = -\gamma(j,i), \qquad i,j \in \mathcal{S},$$

or equivalently, Γ is skew-symmetric:

$$\Gamma^T = -\Gamma.$$

$\gamma(i,j)$ can be written in a more convenient form as shown below. First, for any $T > T_{ij}^*$, we have

$$\begin{aligned} &\int_0^T [f(\widetilde{X}_t) - f(X_t)]dt \\ &= \int_0^{T_{ij}^*} [f(\widetilde{X}_t) - f(X_t)]dt + \int_{T_{ij}^*}^T [f(\widetilde{X}_t) - f(X_t)]dt. \end{aligned}$$

Next, because $\widetilde{X}_{T_{ij}} = X_{T_{ij}}$, by the strong Markov property, we have

$$\lim_{T\to\infty} E\left\{ \int_{T_{ij}^*}^T [f(\widetilde{X}_t) - f(X_t)]dt \,\middle|\, X_0 = i, \widetilde{X}_0 = j \right\} = 0.$$

Thus,

$$\begin{aligned} &\lim_{T\to\infty} E\left\{ \int_0^T [f(\widetilde{X}_t) - f(X_t)]dt \,\middle|\, X_0 = i, \widetilde{X}_0 = j \right\} \\ &= E\left\{ \int_0^{T_{ij}^*} [f(\widetilde{X}_t) - f(X_t)]dt \,\middle|\, X_0 = i, \widetilde{X}_0 = j \right\}, \end{aligned}$$

and from (2.59), we have (cf. (2.6))

$$\gamma(i,j) = \lim_{T\to\infty} E\left\{ \left[\int_0^T f(\widetilde{X}_t)dt - \int_0^T f(X_t)dt \right] \,\middle|\, X_0 = i, \widetilde{X}_0 = j \right\}, \quad i,j \in \mathcal{S}. \tag{2.60}$$

A rigorous proof of (2.60) involves proving the exchangeability of the order of $\lim_{T\to\infty}$ and "E", which follows from the dominated convergence theorem and the finiteness of f, see [62]. Equation (2.60) indicates that $\gamma(i,j)$ measures the "infinitesimal" effect of a perturbation ("jump") from state i to state j on the long-run average reward.

In addition to (2.59) and (2.60), we have another formula that is similar to (2.17) for Markov chains. On the sample path of a Markov process $\boldsymbol{X}$ starting with $X_0 = j$, we define its first passage time to state i as $T(i|j) = \inf\{t : t \geq 0, X_t = i | X_0 = j\}$. Then,

$$\gamma(i,j) = E\left\{ \int_0^{T(i|j)} [f(X_t) - \eta] dt \middle| X_0 = j \right\}. \tag{2.61}$$

An intuitive explanation is similar to Figure 2.8 for (2.17).

For ergodic Markov processes, the PRF matrix Γ satisfies the following *PRF equation*:

$$B\Gamma + \Gamma B^T = -F, \tag{2.62}$$

where $F = ef^T - fe^T$.

Proof. On a Markov process $\boldsymbol{X}$ with $X_0 = i$, we define $p_t(k|i) = \mathcal{P}(X_t = k|X_0 = i)$ and $P_t = [p_t(k|i)]_{i,k\in\mathcal{S}}$. Then, (A.14) gives us

$$P_t = \exp(Bt) = \sum_{n=0}^{\infty} \frac{1}{n!}(Bt)^n, \qquad B^0 = I.$$

It follows that $E[f(X_t)|X_0 = i] = \sum_{k\in\mathcal{S}} p_t(k|i) f(k)$ is the ith entry of $[\exp(Bt)]f$. Let $\widetilde{\boldsymbol{X}}$ be another independent Markov process starting from $\widetilde{X}_0 = j$ and define

$$\begin{aligned}\gamma_T(i,j) &= E\left\{ \int_0^T [f(\widetilde{X}_t) - f(X_t)] dt \middle| X_0 = i, \widetilde{X}_0 = j \right\} \\ &= \int_0^T \left\{ E[f(\widetilde{X}_t)|\widetilde{X}_0 = j] - E[f(X_t)|X_0 = i] \right\} dt, \end{aligned} \tag{2.63}$$

and $\Gamma_T = [\gamma_T(i,j)]_{i,j=1}^S$. Then, from (2.60),

$$\Gamma = \lim_{T\to\infty} \Gamma_T.$$

The integrand on the right-hand side of (2.63) equals the difference between the jth and ith entries of $[\exp(Bt)]f$. Therefore,

$$\Gamma_T = \int_0^T \left\{ ef^T [\exp(Bt)]^T - [\exp(Bt)] fe^T \right\} dt.$$

Using $Be = 0$ and $[\exp(Bt)]B = B[\exp(Bt)]$, we obtain

$$\begin{aligned} & B\Gamma_T + \Gamma_T B^T \\ &= \int_0^T \left\{ ef^T[\exp(Bt)]^T B^T - B[\exp(Bt)]fe^T \right\} dt \\ &= ef^T \left[\int_0^T [\exp(Bt)]Bdt \right]^T - \left[\int_0^T [\exp(Bt)]Bdt \right] fe^T \\ &= ef^T[\exp(BT) - \exp(0)]^T - [\exp(BT) - \exp(0)]fe^T, \end{aligned} \tag{2.64}$$

where the variable 0 in exp(0) denotes a matrix whose elements are all zeros. Therefore, $\exp(0) = I$. For ergodic Markov processes, we have $\lim_{T\to\infty} p_T(j|i) = \pi(j)$; thus, $\lim_{T\to\infty} \exp(BT) = \lim_{T\to\infty} P_T = e\pi$. Furthermore, $ef^T(e\pi)^T = (\pi f)ee^T = e\pi fe^T$. Letting $T \to \infty$ in (2.64), we obtain the PRF equation (2.62). □

If F is a Hermitian matrix, then (2.62) is the continuous-time version of the Lyapunov equation [162, 174]. However, the continuous-time PRF equation (2.62) is different from the Lyapunov equation because F here is a skew-symmetric matrix, $F^T = -F$.

Next, it is easy to see that the solution to (2.62) with the form of (2.65), specified below, is unique. Suppose that there are two such solutions to (2.62) denoted as $\Gamma_1 = eg_1^T - g_1e^T$ and $\Gamma_2 = eg_2^T - g_2e^T$. Let $W = \Gamma_1 - \Gamma_2 = ew^T - we^T$, with $w = g_1 - g_2$. Then $BW + WB^T = 0$. Because $Be = 0$, we have $ew^T B^T - Bwe^T = 0$. Multiplying both sides of this equation on the left by the group inverse $B^\#$ and using $B^\# B = I - e\pi$ and $B^\# e = 0$, we have $(I - e\pi)we^T = 0$. Therefore,

$$we^T = e\pi we^T = (\pi w)ee^T,$$

where πw is a constant. From this, we have $W = (we^T)^T - we^T = 0$, i.e., $\Gamma_1 = \Gamma_2$.

Performance Potentials

From (2.60), we have

$$\gamma(i,j) = \gamma(i,k) + \gamma(k,j), \qquad i,j,k \in \mathcal{S}.$$

Similar to the sensitivity analysis of Markov chains, we can define *performance potentials* $g(i)$, $i \in \mathcal{S}$, as follows:

$$\gamma(i,j) = g(j) - g(i), \qquad \text{for all } i,j \in \mathcal{S},$$

or, equivalently,

$$\Gamma = eg^T - ge^T, \tag{2.65}$$

where $g = (g(1), \ldots, g(S))^T$ is called a *potential vector*.

Substituting (2.65) into (2.62), we get $e(Bg + f)^T = (Bg + f)e^T$. Thus, $Bg + f = ce$, with c being a constant. Because $\pi B = 0$, we get $c = \pi f = \eta$. Thus, the performance potentials satisfy the following *Poisson equation*:

$$\boxed{Bg = -f + \eta e. \qquad (2.66)}$$

Again, its solution is only up to an additive constant: if g is a solution to (2.66), so is $g + ce$ for any constant c.

For ergodic Markov processes, the group inverse of B is defined as $B^{\#} = (B - e\pi)^{-1} + e\pi$ [202] (cf. (2.48)). We have

$$BB^{\#} = B^{\#}B = I - e\pi.$$

By multiplying both sides of (2.66) on the left by $B^{\#}$, we obtain the general form of its solution

$$g = -B^{\#}f + ce,$$

where $c = \pi g$, which may be any constant. In particular, we can choose a solution that satisfies $c = \pi g = \eta$. In this case, the Poisson equation (2.66) becomes

$$(B - e\pi)g = -f,$$

and its solution is

$$g = -B^{\#}f + \eta e.$$

We may also choose $c = \pi g = 0$. Then,

$$\boxed{g = -B^{\#}f. \qquad (2.67)}$$

If we choose $c = \pi g = -\eta$, then

$$g = -(B + e\pi)^{-1}f = -(B^{\#} + e\pi)f = -B^{\#}f - \eta e.$$

For simplicity, we have used the same notation g to denote different versions of the potentials, which may differ by a constant. We need to keep this in mind to avoid possible confusion.

Now, let us develop a sample-path-based explanation for $B^{\#}$ and g. First, we have

$$\int_0^\infty B[\exp(Bt)]dt = \int_0^\infty [\exp(Bt)]Bdt = -(I - e\pi).$$

From this, using $Be = \pi B = 0$, we get

$$B\left\{\int_0^\infty [\exp(Bt) - e\pi]dt\right\} = \left\{\int_0^\infty [\exp(Bt) - e\pi]dt\right\}B = -(I - e\pi). \quad (2.68)$$

Furthermore, we can easily prove that

$$\pi\left\{\int_0^\infty [\exp(Bt) - e\pi]dt\right\} = \left\{\int_0^\infty [\exp(Bt) - e\pi]dt\right\}e = 0.$$

By multiplying both sides of (2.68) on the left by $B^{\#}$, we obtain

$$\begin{aligned} B^{\#} &= -\int_0^\infty [\exp(Bt) - e\pi]dt \\ &= -\lim_{T\to\infty}\left[\int_0^T \exp(Bt)dt - Te\pi\right]. \end{aligned} \tag{2.69}$$

From (2.69) and (2.67), and using $P_t = \exp(Bt)$, we get

$$\boxed{g(i) = \lim_{T\to\infty} E\left\{\int_0^T [f(X_t) - \eta]dt \middle| X_0 = i\right\}.}$$

This is the sample path explanation of the potential $g(i)$ (cf. (2.16)). This is also consistent with (2.60).

In modern Markov theory [87], the *α-potential* of a function f is defined as

$$g^{(f)}(i) = E\left\{\int_0^\infty [\exp(-\alpha t)]f(X_t)dt \middle| X_0 = i\right\}, \qquad \alpha > 0.$$

Thus, our definition of the potential can be viewed as an extension of the classical α-potential to the case of $\alpha = 0$. To keep the integral finite at $\alpha = 0$, a constant term η is subtracted from the integrand (see (2.46) for the discussion of the discrete-time version).

Performance Derivatives

With the aforementioned results, the performance derivative formulas can be easily derived. Let B and B' be two infinitesimal generators on the same state space $\mathcal{S}$. Suppose that B changes to another infinitesimal generator $B_\delta = [b_\delta(i,j)] = B + \delta\varDelta B$, with $\delta > 0$ being a small real number, $\varDelta B = B' - B = [\varDelta b(i,j)]$. We have $\varDelta Be = 0$. Let $\boldsymbol{X}_\delta$ be the Markov process with infinitesimal generator B_δ. We assume that $\boldsymbol{X}_\delta$ is also irreducible. Let π_δ be the vector of the steady-state probabilities of $\boldsymbol{X}_\delta$. The average reward of $\boldsymbol{X}_\delta$ is $\eta_\delta = \eta + \varDelta\eta_\delta$. The performance derivative along the direction of $\varDelta B$ is $\frac{d\eta_\delta}{d\delta}\big|_{\delta=0} = \lim_{\delta\to 0}\frac{\eta_\delta - \eta}{\delta}$. With this notation, we have $\frac{dB_\delta}{d\delta} = \varDelta B$.

Taking derivatives of both sides of $\pi_\delta B_\delta = 0$ at $\delta = 0$, we get

$$\left.\frac{d\pi_\delta}{d\delta}\right|_{\delta=0} B = -\pi\frac{dB_\delta}{d\delta} = -\pi(\varDelta B).$$

By multiplying both sides of this equation on the right by $B^{\#}$ and using $BB^{\#} = I - e\pi$ and $\frac{d\pi_\delta}{d\delta} e = 0$, we obtain

$$\left.\frac{d\pi_\delta}{d\delta}\right|_{\delta=0} = -\pi(\Delta B)B^{\#}.$$

Therefore,

$$\left.\frac{d\eta_\delta}{d\delta}\right|_{\delta=0} = -\pi(\Delta B)B^{\#}f = \pi(\Delta B)g.$$

Next, by multiplying both sides of (2.62) on the right by π^T and using $\pi B = 0$ and $\pi e = 1$, we have

$$B\Gamma\pi^T = fe^T\pi^T - ef^T\pi^T = (I - e\pi)f.$$

That is, $B\Gamma\pi^T = BB^{\#}f$. By multiplying both sides of this equation on the left by $B^{\#}$, we get $(I - e\pi)\Gamma\pi^T = (I - e\pi)B^{\#}f$. Using $\pi B^{\#} = 0$ and $\pi\Gamma\pi^T = \pi(eg^T - ge^T)\pi^T = 0$, we obtain

$$B^{\#}f = \Gamma\pi^T.$$

This leads to the performance derivative formula in terms of Γ:

$$\left.\frac{d\eta_\delta}{d\delta}\right|_{\delta=0} = -\pi(\Delta B)\Gamma\pi^T.$$

If, in addition to the changes in B, the reward function f also changes to $f_\delta = f + \delta\Delta f$, we have

$$\boxed{\left.\frac{d\eta_\delta}{d\delta}\right|_{\delta=0} = \pi[(\Delta B)g + \Delta f].}$$

The higher-order derivatives can be derived in a way similar to the Markov chains:

$$\boxed{\left.\frac{d^n\eta_\delta}{d\delta^n}\right|_{\delta=0} = n!\pi\left\{[(\Delta B)(-B^{\#})]^{n-1}[(\Delta B)(-B^{\#})f + \Delta f]\right\}. \qquad (2.70)}$$

In addition, we have the following MacLaurin expansion:

$$\eta_\delta = \eta + \pi\sum_{k=1}^{n}[\delta(\Delta B)(-B^{\#})]^{k-1}[(\Delta B)(-B^{\#})f + \Delta f]\delta$$
$$+ \pi_\delta[\delta(\Delta B)(-B^{\#})]^n[(\Delta B)(-B^{\#})f + \Delta f]\delta.$$

When $\Delta f = 0$, this becomes

$$\boxed{\eta_\delta = \pi \sum_{k=0}^{n} [\delta(\Delta B)(-B^{\#})]^k f + \pi_\delta [\delta(\Delta B)(-B^{\#})]^{n+1} f.}$$

Thus, we can use $\pi \sum_{k=0}^{n} [\delta(\Delta B)(-B^{\#})]^k f$ to estimate η_δ, and the error in the estimation is $\pi_\delta [\delta(\Delta B)(-B^{\#})]^{n+1} f$. All the items in π and $B^{\#}$ can be estimated on a sample path of the Markov process with infinitesimal generator B, see Problem 3.18.

2.3 Performance Sensitivities of Semi-Markov Processes*

In this section, we extend the above PA results to (continuous-time) semi-Markov processes (SMPs). The previous results on PA of Markov processes become special cases. This section is based on [57], and we only study the long-run average-reward problem (for extensions to the discounted-reward problem, see [57]).

2.3.1 Fundamentals for Semi-Markov Processes*

We study a semi-Markov process $\boldsymbol{X} = \{X_t, t \geq 0\}$ defined on a finite state space $\mathcal{S} = \{1, 2, \ldots, S\}$. Let $T_0, T_1, \ldots, T_l, \ldots$, with $T_0 = 0$, be the transition epoches. The process is right continuous so the state at each transition epoch is the state after the transition. Let $X_l = X_{T_l}$, $l = 0, 1, 2, \ldots$. Then, $\{X_0, X_1, \ldots\}$ is the embedded Markov chain. The interval $[T_l, T_{l+1})$ is called a *period* and its length is called the *sojourn time* in state X_l.

The Embedded Chain and the Sojourn Time

The semi-Markov kernel [87] is defined as

$$p(j;t|i) := \mathcal{P}\left(X_{l+1} = j, T_{l+1} - T_l \leq t | X_l = i\right),$$

which we assume does not depend on l (time-homogenous). Set

$$p(t|i) := \sum_{j \in \mathcal{S}} p(j;t|i) = \mathcal{P}(T_{l+1} - T_l \leq t | X_l = i),$$

$$h(t|i) := 1 - p(t|i),$$

$$p(j|i) := \lim_{t \to \infty} p(j;t|i) = \mathcal{P}(X_{l+1} = j | X_l = i),$$

and

$$p(t|i,j) := \frac{p(j;t|i)}{p(j|i)} = \mathcal{P}(T_{l+1} - T_l \le t | X_l = i, X_{l+1} = j).$$

Normally, $p(i|i) = 0$, for all $i \in \mathcal{S}$. But, in general, we may allow the process to move from a state to itself at the transition epoches; in such a case, $p(i|i)$ may be nonzero and our results still hold. However, a transition from a state to the same state cannot be determined by observing only the system states of a semi-Markov process.

The matrix $[p(j|i)]$ is the transition probability matrix of the embedded Markov chain. We assume that this matrix is irreducible and aperiodic [20]. Let

$$m(i) = \int_0^\infty sp(ds|i) = E[T_{l+1} - T_l | X_l = i]$$

be the mean of the sojourn time in state i. We also assume that $m(i) < \infty$ for all $i \in \mathcal{S}$. Under these assumptions, the semi-Markov process is irreducible and aperiodic and hence ergodic. Define the *hazard rates* as

$$r(t|i) = \frac{\frac{d}{dt} p(t|i)}{h(t|i)},$$

and

$$r(j;t|i) = \frac{\frac{d}{dt} p(j;t|i)}{h(t|i)}.$$

The latter is the rate at which the process moves from i to j in $[t, t+dt)$ given that the process does not move out from state i in $[0, t)$.

The Equivalent Infinitesimal Generator

Let $p_t(j|i) = \mathcal{P}(X_t = j | X_0 = i)$. By the total probability theorem, we can easily derive

$$p_{t+\Delta t}(j|i) = \sum_{k \in \mathcal{S}} p_t(k|i) \int_0^\infty \widetilde{p}_t(s|k)\{I_j(k)[1 - r(s|k)\Delta t] + r(j;s|k)\Delta t\} ds, \tag{2.71}$$

where $I_j(k) = 1$ if $k = j$, $I_j(k) = 0$ if $k \neq j$; $\widetilde{p}_t(s|k)ds$ is defined as the probability that, given that the state at time t is k, the process has been in state k for a period of s to $s + ds$. This probability depends on k and therefore may depend on the initial state. Precisely, let l_t be the integer such that $T_{l_t} \le t < T_{l_t+1}$. Then,

$$\widetilde{p}_t(s|k)ds = \mathcal{P}(s \le t - T_{l_t} < s + ds | X_t = k). \tag{2.72}$$

It is proved at the end of this subsection that

$$\lim_{t\to\infty} \widetilde{p}_t(s|k) = \frac{h(s|k)}{m(k)}. \tag{2.73}$$

Now, set $\Delta t \to 0$ in (2.71) and we obtain

$$\frac{dp_t(j|i)}{dt} = -\sum_{k \in \mathcal{S}} p_t(k|i) \int_0^\infty \{\widetilde{p}_t(s|k)[I_j(k) r(s|k) - r(j;s|k)]\} ds. \tag{2.74}$$

Since the semi-Markov process is ergodic, when $t \to \infty$, we have $p_t(j|i) \to \pi(j)$ [87] and $\frac{dp_t(j|i)}{dt} \to 0$, where $\pi(j)$ is the steady-state probability of j. Letting $t \to \infty$ on both sides of (2.74) and using (2.73), we get

$$\begin{aligned} 0 &= -\sum_{k \in \mathcal{S}} \pi(k) \int_0^\infty \frac{1}{m(k)} \left\{ I_j(k) \frac{d}{ds}[p(s|k)] - \frac{d}{ds}[p(j;s|k)] \right\} ds \\ &= -\sum_{k \in \mathcal{S}} \pi(k) \left\{ \frac{1}{m(k)} [I_j(k) - p(j|k)] \right\} \\ &= -\sum_{k \in \mathcal{S}} \pi(k) \{\lambda(k)[I_j(k) - p(j|k)]\}, \end{aligned}$$

where we define

$$\lambda(k) := \frac{1}{m(k)}.$$

Finally, we have

$$\sum_{k \in \mathcal{S}} \pi(k) b(k,j) = 0, \qquad \text{for all } j \in \mathcal{S},$$

where we define

$$\boxed{b(k,j) = -\lambda(k)[I_j(k) - p(j|k)]. \qquad (2.75)}$$

In matrix form, we can write

$$\pi B = 0, \tag{2.76}$$

where $\pi = (\pi(1), \ldots, \pi(S))$ is the steady-state probability vector and B is a matrix with elements $b(k,j)$. In addition, we can easily verify that

$$Be = 0.$$

Equation (2.76) is exactly the same as the Markov process with B as its infinitesimal generator. Therefore, B in (2.76) is *the equivalent infinitesimal generator* for a semi-Markov process. Note that B depends only on $m(i)$ and $p(j|i)$, $i, j \in \mathcal{S}$. This implies that the steady-state probability is insensitive to the high-order statistics of the sojourn times in any state, and it is independent of whether the sojourn time in state i depends on j, the state it moves into from i.

The Steady-State Probability

We will study the general case where the reward function depends not only on the current state but also on the next state that the semi-Markov process moves into. To this end, for any time $t \in [T_l, T_{l+1})$, we denote $Y_t = X_{l+1}$, and study the process $\{(X_t, Y_t), t \geq 0\}$. Because the process $\{Y_t, t \geq 0\}$ is completely determined by the process $\{X_t, t \geq 0\}$, for notational simplicity, we still denote the process $\{(X_t, Y_t), t \geq 0\}$ as

$$\boldsymbol{X} = \{(X_t, Y_t), t \geq 0\}. \tag{2.77}$$

Let $\pi(i,j)$ be the steady-state probability of $(X_t, Y_t) = (i,j)$ and $\pi(j|i)$ be the steady-state conditional probability of $Y_t = j$ given that $X_t = i$, i.e., $\pi(j|i) = \lim_{t\to\infty} P(Y_t = j|X_t = i)$. (This is different from $\lim_{l\to\infty} P(X_{l+1} = j|X_l = i)$, which is the steady-state conditional probability of the embedded Markov chain.)

Define

$$m(i,j) = \int_0^\infty sp(ds|i,j) = E[T_{l+1} - T_l|X_l = i, X_{l+1} = j].$$

Then, we have

$$m(i) = \sum_{j\in\mathcal{S}} p(j|i)m(i,j) = \int_0^\infty sp(ds|i). \tag{2.78}$$

We can prove (see the end of this subsection)

$$\pi(j|i) = \frac{\int_0^\infty sp(j;ds|i)}{\int_0^\infty sp(ds|i)} = \frac{p(j|i)m(i,j)}{m(i)}. \tag{2.79}$$

Therefore,

$$\pi(i,j) = \pi(j|i)\pi(i) = \pi(i)\frac{p(j|i)m(i,j)}{m(i)}, \tag{2.80}$$

where $\pi(i)$, $i \in \mathcal{S}$, can be obtained from (2.76).

Proofs

A. The Proof of (2.73).

Consider an interval $[0, T_L]$, with $L >> 1$. Let $I_k(x) = 1$ if $x = k$ and $I_k(x) = 0$ if $x \neq k$; and $I(*)$ be an indicator function, i.e., $I(*) = 1$ if the expression in the brackets holds, $I(*) = 0$ otherwise. Let l_t be the integer such that $T_{l_t} \leq t < T_{l_t+1}$. From (2.72), by ergodicity, we have

$$\lim_{t\to\infty} \widetilde{p}_t(s|k)ds = \lim_{T_L\to\infty} \frac{\int_0^{T_L} I(s \leq t - T_{l_t} < s + ds)I_k(X_t)dt}{\int_0^{T_L} I_k(X_t)dt}. \tag{2.81}$$

Let N_k be the number of periods in $[0, T_L]$ in which $X_t = k$. We have

$$\lim_{T_L \to \infty} \frac{1}{N_k} \int_0^{T_L} I_k(X_t) dt = \int_0^\infty sp(ds|k). \tag{2.82}$$

Next, we observe that, for a fixed $s > 0$, $\int_0^{T_L} I(s \leq t - T_{l_t}) I_k(X_t) dt$ is the total length of the time period in $[0, T_L]$ in which $s \leq t - T_{l_t}$ and $X_t = k$. Furthermore, among the N_k periods, roughly $N_k p(d\tau|k)$ periods terminate with a length of τ to $\tau + d\tau$. For any $s < \tau$, in each of such periods, the length of time in which $s \leq t - T_{l_t}$ is $\tau - s$. Thus,

$$\int_0^{T_L} I(s \leq t - T_{l_t}) I_k(X_t) dt \approx N_k \int_s^\infty (\tau - s) p(d\tau|k),$$

or

$$\lim_{T_L \to \infty} \frac{1}{N_k} \int_0^{T_L} I(s \leq t - T_{l_t}) I_k(X_t) dt = \int_s^\infty (\tau - s) p(d\tau|k).$$

Therefore,

$$\begin{aligned}
&\lim_{T_L \to \infty} \frac{1}{N_k} \int_0^{T_L} I(s \leq t - T_{l_t} < s + ds) I_k(X_t) dt \\
&= - \lim_{T_L \to \infty} \frac{1}{N_k} \left[\int_0^{T_L} I(s + ds \leq t - T_{l_t}) I_k(X_t) dt \right. \\
&\qquad \left. - \int_0^{T_L} I(s \leq t - T_{l_t}) I_k(X_t) dt \right] \\
&= - \frac{d}{ds} \left[\int_s^\infty (\tau - s) p(d\tau|k) \right] ds = [1 - p(s|k)] ds = h(s|k) ds.
\end{aligned} \tag{2.83}$$

From (2.81), (2.82), and (2.83), we get

$$\lim_{t \to \infty} \widetilde{p}_t(s|k) ds = \frac{h(s|k) ds}{\int_0^\infty sp(ds|k)}.$$

Therefore, (2.73) holds. □

B. The Proof of (2.79).

Consider a time interval $[0, T_L]$, with $L >> 1$. Let N_i be the number of periods in $[0, T_L]$ in which $X_t = i$. Then,

$$\lim_{T_L \to \infty} \frac{1}{N_i} \int_0^{T_L} I_i(X_t) dt = \int_0^\infty sp(ds|i).$$

Let $I_{i,j}(x, y) = 1$ if $x = i$ and $y = j$, and $I_{i,j}(x, y) = 0$ otherwise. We have

$$\lim_{T_L \to \infty} \frac{1}{N_i} \int_0^{T_L} I_{i,j}(X_t, Y_t) dt = \int_0^\infty sp(j; ds|i).$$

Thus, we have

$$\pi(j|i) = \lim_{T_L \to \infty} \frac{\int_0^{T_L} I_{i,j}(X_t, Y_t)dt}{\int_0^{T_L} I_i(X_t)dt} = \frac{\int_0^\infty sp(j;ds|i)}{\int_0^\infty sp(ds|i)} = \frac{p(j|i)m(i,j)}{m(i)}.$$

Therefore, (2.79) holds. □

2.3.2 Performance Sensitivity Formulas*

Consider a semi-Markov process $\boldsymbol{X} = \{(X_t, Y_t), t \geq 0\}$ (see (2.77)) starting from a *transition epoch* $T_0 = 0$ and an initial state $X_0 = j$. We define the reward function as $f(i,j)$, $i, j \in \mathcal{S}$, where $f : \mathcal{S} \times \mathcal{S} \to \mathcal{R}$. The long-run average reward is

$$\eta = \lim_{T \to \infty} \frac{1}{T} E\left[\int_0^T f(X_t, Y_t)dt \middle| X_0 = j\right],$$

which does not depend on j because $\boldsymbol{X}$ is ergodic.

The Perturbation Realization Factor

On $\boldsymbol{X}$ with $T_0 = 0$ and $X_0 = j$, denote the instant at which the process moves into state i for the first time as

$$T(i|j) = \inf\{t : t \geq 0,\ X_t = i | X_0 = j\}.$$

Following the same approach as for the PA of Markov processes (2.61), we define the *perturbation realization factors* as (the only difference is that $T_0 = 0$ must be a transition epoch in the semi-Markov case):

$$\gamma(i,j) = E\left\{ \int_0^{T(i|j)} [f(X_t, Y_t) - \eta]dt \middle| X_0 = j \right\}. \tag{2.84}$$

Define $\varGamma = [\gamma(i,j)]_{i,j=1}^S$.

From (2.80) and by ergodicity, we have

$$\eta = \sum_{i,j \in \mathcal{S}} \pi(i,j)f(i,j) = \sum_{i \in \mathcal{S}} \pi(i)f(i) = \pi f,$$

where $f = (f(1), f(2), \ldots, f(S))^T$, and (for simplicity, we use "f" for both $f(i)$ and $f(i,j)$)

$$f(i) = \frac{\sum_{j \in \mathcal{S}} p(j|i)f(i,j)m(i,j)}{m(i)}. \tag{2.85}$$

From (2.84), we have

$$\begin{aligned}\gamma(i,j) &= E\left\{\int_0^{T_1}[f(X_t,Y_t)-\eta]dt\middle| X_0=j\right\}\\ &\qquad + E\left\{\int_{T_1}^{T(i|j)}[f(X_t,Y_t)-\eta]dt\middle| X_0=j\right\}\\ &=\sum_{k\in\mathcal{S}}p(k|j)\left\{E\left\{\int_0^{T_1}[f(X_0,Y_0)-\eta]dt\middle| X_0=j,X_1=k\right\}\right.\\ &\qquad \left.+E\left\{\int_{T_1}^{T(i|j)}[f(X_t,Y_t)-\eta]dt\middle| X_0=j,X_1=k\right\}\right\}\\ &=\sum_{k\in\mathcal{S}}p(k|j)\Bigg\{[f(j,k)-\eta]E[T_1|X_0=j,X_1=k]\\ &\qquad \left.+E\left\{\int_{T_1}^{T(i|k)}[f(X_t,Y_t)-\eta]dt\middle| X_1=k\right\}\right\}\\ &=\sum_{k\in\mathcal{S}}p(k|j)\Bigg\{[f(j,k)-\eta]m(j,k)\\ &\qquad \left.+ E\left\{\int_{T_1}^{T(i|k)}[f(X_t,Y_t)-\eta]dt\middle| X_1=k\right\}\right\}.\end{aligned}$$

From (2.78) and (2.85), the aforementioned equation leads to

$$\gamma(i,j)=m(j)[f(j)-\eta]+\sum_{k\in\mathcal{S}}p(k|j)\gamma(i,k),$$

or, equivalently,

$$\begin{aligned}-[f(j)-\eta] &= \sum_{k\in\mathcal{S}}\{-\lambda(j)[I_j(k)-p(k|j)]\gamma(i,k)\}\\ &= \sum_{k\in\mathcal{E}}[b(j,k)\gamma(i,k)].\end{aligned}$$

In matrix form, this is

$$\Gamma B^T=-(ef^T-\eta ee^T). \tag{2.86}$$

Next, on the process $\boldsymbol{X}$, with $T_0=0$ being a transition epoch and $X_0=j$, for any state $i\in\mathcal{S}$ we define two sequences $u_0,u_1,\ldots,$ and $v_0,v_1,\ldots,$ as follows:

$$u_0=T_0=0, \tag{2.87}$$

$$v_n=\inf\{t\geq u_n, X_t=i\},$$

and

$$u_{n+1} = \inf\{t \geq v_n, X_t = j\}, \tag{2.88}$$

i.e., v_n is the first time when the process reaches i after u_n, and u_{n+1} is the first time when the process reaches j after v_n, $n = 0, 1, \ldots$. Apparently, $u_0, u_1, \ldots$ are stopping times and X_t is a regenerative process with $\{u_n, n = 0, 1, \ldots\}$ as its associated renewal process. By the theory of regenerative processes [87], we have

$$\eta = \frac{E[\int_{u_0}^{u_1} f(X_t, Y_t)dt]}{E[u_1 - u_0]} = \frac{E[\int_0^{v_0} f(X_t, Y_t)dt] + E[\int_{v_0}^{u_1} f(X_t, Y_t)dt]}{E[v_0] + E[u_1 - v_0]}.$$

Thus,

$$E\left\{\int_0^{v_0} [f(X_t, Y_t) - \eta]dt\right\} + E\left\{\int_{v_0}^{u_1} [f(X_t, Y_t) - \eta]dt\right\} = 0.$$

By the definition of u_0, v_0 and u_1, the above equation is

$$\gamma(i, j) + \gamma(j, i) = 0;$$

therefore, the matrix $\varGamma$ is skew-symmetric

$$\varGamma^T = -\varGamma.$$

Taking the transpose of (2.86), we get

$$-B\varGamma = -(fe^T - \eta ee^T).$$

From the above equation and (2.86), $\varGamma$ satisfies the following *PRF equation*

$$B\varGamma + \varGamma B^T = -F,$$

where $F = ef^T - fe^T$.

Performance Potentials

Similar to Equations (2.87) to (2.88), for any three states i, j, k, we define three sequences $u_0, u_1, \ldots$; $v_0, v_1, \ldots$; and $w_0, w_1, \ldots$ as follows. $u_0 = T_0 = 0$, $X_0 = j$, $v_n = \inf\{t \geq u_n, X_t = i\}$, $w_n = \inf\{t \geq v_n, X_t = k\}$, and $u_{n+1} = \inf\{t \geq w_n, X_t = j\}$. By a similar approach, we can prove that

$$\gamma(i, j) + \gamma(j, k) + \gamma(k, i) = 0.$$

In general, we can prove that, for any closed circle $i_1 \to i_2 \to \cdots \to i_n \to i_1$ in the state space, we have

$$\gamma(i_1, i_2) + \gamma(i_2, i_3) + \cdots + \gamma(i_{n-1}, i_n) + \gamma(i_n, i_1) = 0.$$

This is similar to the conservation law of potential energy in physics. Therefore, we can define a performance potential $g(i)$ in any state and write $\gamma(i,j) = g(j) - g(i)$ and

$$\Gamma = eg^T - ge^T, \tag{2.89}$$

where $g = (g(1), \dots, g(S))^T$. By substituting (2.89) into (2.86), we get the *Poisson equation*:

$$Bg = -f + \eta e.$$

Similar to the Markov process case, we have different versions of g, which differ by a constant vector ce. For example, when $\pi g = \eta$, we have

$$(B - e\pi)g = -f, \tag{2.90}$$

and when $\pi g = 0$, we have

$$g = -B^{\#} f.$$

Finally, a Markov process with transition rates $\lambda(i)$ and transition probabilities $p(j|i)$ can be viewed as a semi-Markov process whose kernel is $p(j;t|i) = p(j|i)\{1 - \exp[-\lambda(i)t]\}$. With this special kernel, we have

$$m(i,j) = m(i) = \frac{1}{\lambda(i)},$$

$$\pi(i,j) = \pi(i)p(j|i),$$

and

$$f(i) = \sum_{j=1}^{S} p(j|i)f(i,j).$$

The results in this section become the same as those in Section 2.2 for Markov processes.

Performance Sensitivity Formulas

We have shown that with properly defined g and B, the Poisson equation and PRF equation hold for potentials and perturbation realization factor matrices, respectively, for semi-Markov processes. Thus, performance sensitivity formulas can be derived in a way similar to Markov processes, and the results are briefly stated here.

First, for two semi-Markov processes with B', η', f' and B, η, f, by multiplying both sides of (2.90) on the left by π' and using $\pi' B' = 0$ and $\pi g = \eta$, we get

$$\begin{aligned} \eta' - \eta &= \pi'[(B' - B)g + (f' - f)] \\ &= \pi'[(B'g + f') - (Bg + f)]. \end{aligned} \tag{2.91}$$

As we shall see in Chapter 4, this equation serves as a foundation for semi-Markov decision processes. As shown in Chapter 4, policy iteration for semi-Markov processes can be derived from (2.91).

Next, suppose that B changes to $B_\delta = B + \delta\Delta B$, with $\Delta B = B' - B$, and f changes to $f_\delta = f + \delta\Delta f$, with $\Delta f = f' - f$. We have $\Delta Be = 0$. ΔB can be determined by the changes in the characteristics of the semi-Markov process. For example, if $\lambda(i) = 1/m(i)$ changes to $\lambda(i)+(\Delta\lambda)\delta$, $i = 1, 2, \ldots, S$, $\Delta\lambda > 0$, then, according to (2.75), $b(i,j)$ changes to $b(i,j) - \delta(\Delta\lambda)[I_j(i) - p(j|i)]$; i.e., $\Delta B = -\Delta\lambda(I - P)$, $P = [p(j|i)]$; on the other hand, if P changes to $P + \Delta P$, then $\Delta b(i,j) = \lambda(i)[\Delta P(j|i)]$, $i, j = 1, 2, \ldots, S$. Denote the average reward of the semi-Markov system with B_δ and f_δ as η_δ. We can easily obtain

$$\boxed{\begin{aligned} \left.\frac{d\eta_\delta}{d\delta}\right|_{\delta=0} &= \pi[-(\Delta B)B^{\#}f + \Delta f] \\ &= \pi[(\Delta B)\Gamma^T\pi^T + \Delta f]. \end{aligned}}$$

Sample-path-based expressions for g and Γ can be derived. From (2.84), with a similar reasoning as in (2.18), we have

$$\begin{aligned} \gamma(i,j) = \lim_{T\to\infty} \Bigg\{ & E\Big\{ \int_0^T [f(\widetilde{X}_t, \widetilde{Y}_t) - \eta]dt \Big| \widetilde{X}_0 = j \Big\} \\ & - E\Big\{ \int_0^T [f(X_t, Y_t) - \eta]dt \Big| X_0 = i \Big\} \Bigg\}, \end{aligned}$$

where $\widetilde{\boldsymbol{X}}$ and $\boldsymbol{X}$ have the same kernel; they are independent but start from two different initial states, $\widetilde{X}_0 = j$ and $X_0 = i$, respectively, with $T_0 = \widetilde{T}_0 = 0$ being a transition epoch for both $\widetilde{\boldsymbol{X}}$ and $\boldsymbol{X}$. From this equation, we have

$$\boxed{g(j) = \lim_{T\to\infty} E\Big\{ \int_0^T [f(X_t, Y_t) - \eta]dt \Big| X_0 = j \Big\}. \qquad (2.92)}$$

This is the same as in the Markov process case, except that the integral starts with a transition epoch. The convergence of the right-hand side of (2.92) can be easily verified by, e.g., using the embedded Markov chain model.

With the equivalent infinitesimal generator, the high-order derivatives are the same as those for the Markov chains (2.70). Again, all the items in π and $B^{\#}$ can be estimated on a sample path of the semi-Markov process with B; see Problem 3.18.

Example 2.3. Consider a communication line (or a switch, a router, etc.) at which packets arrive in a Poisson process with a rate of λ packets per second. The packet length is assumed to have a general probability distribution function $\varPhi(x)$; the unit of the length is bit per packet. For each packet, the system manager can choose a transmission rate of θ bits per second. Thus, the transmission time for each packet has a distribution function $\widetilde{\varPhi}(\tau) = \mathcal{P}(t \le \tau) = \mathcal{P}(x \le \theta\tau) = \varPhi(\theta\tau)$. In a real system, θ takes a discrete value determined by the number of channels; each channel has a fixed amount of bandwidth. Thus, we can view θ as an action and denote the action space as $\{\theta_1, \theta_2, \ldots, \theta_K\}$, with $\theta_k = k\mu$, $k = 1, 2, \ldots, K$, where μ denotes the transmission rate of one channel in bits per second. Of course, in a theoretical study, we can also view θ as a continuous variable.

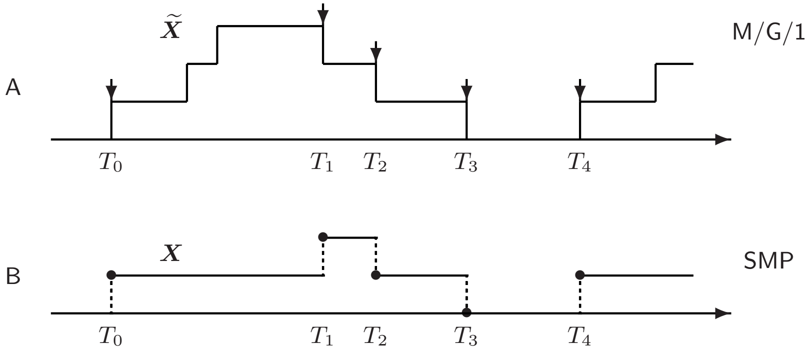


Fig. 2.11. An M/G/1 Queue and the Embedded SMP

The system can be modelled as an M/G/1 queue; the (physical) state at time t is $N(t) = i$ with i being the number of customers (packets) in the queue at time t. Figure 2.11.A illustrates a sample path $\boldsymbol{X} = \{N(t), t \ge 0\}$. For stability, we require that $K\mu > \lambda\bar{x}$, where $\bar{x}$ is the mean length of the packets. The decisions for actions are made at the beginning of the transmission of every packet. Thus, we consider the embedded points consisting of all the service completion times and the arrival times to all the idle periods, denoted as $T_0, T_1, \ldots$. Define $\widetilde{X}_t = N(T_n)$ for $T_n \le t < T_{n+1}$, $n = 0, 1, 2, \ldots$. Then, $\widetilde{\boldsymbol{X}} = \{\widetilde{X}_t, t \ge 0\}$ is a semi-Markov process (SMP). Figure 2.11.B illustrates the embedded SMP corresponding to the sample path in Figure 2.11.A. It is clear that the following equations hold for $\widetilde{\boldsymbol{X}}$:

$$p(1; t|0) = 1 - \exp(-\lambda t),$$

$$p(t|i) = \varPhi(\theta t), \qquad i > 0,$$

and

$$p(j; dt|i) = \mathcal{P}(X_{n+1} = j, t \le T_{n+1} - T_n < t + dt | X_n = i)$$

$$= \left[\frac{(\lambda t)^{j-i+1}}{(j-i+1)!} \exp(-\lambda t) \right] \Phi(\theta dt), \qquad i > 0,\ i-1 \leq j,$$

where the term in the braces is the probability that there are $j-i+1$ arrivals in the period of $[0, t)$.

In the optimization problem, the reward (cost) function usually consists of two parts: the holding cost $f_1(i,j)$ and the bandwidth cost $f_2(\theta)$. That is,

$$f_\theta(i,j) = f_1(i,j) + f_2(\theta).$$

It is well known that, if in an interval $[0,t]$, there are k arrivals from a Poisson process, then these k arrivals uniformly distribute over the period (see, e.g., [169]). Thus, it is reasonable to take the average number of customers in $[0,t]$, $(i+j)/2$, as the holding cost, and we may set

$$f_\theta(i,j) = \kappa_1 \frac{i+j}{2} + \kappa_2 \theta, \qquad \kappa_1 + \kappa_2 = 1,\ 0 < \kappa_1, \kappa_2 < 1,$$

where the first term represents the cost for the average waiting time. The problem is now formulated in a semi-Markov framework and the results developed in this section can be applied. □

Finally, many results about SMPs can be obtained by using the embedded Markov chain method (see, e.g., [243]). It is natural to expect that the sensitivity analysis can also be implemented using this approach. However, compared with the embedded-chain-based approach, the approach presented in this section is more direct and concise and hence the results have a clear intuitive interpretation. In addition, with the embedded approach, the expected values (time and cost) on a period $T_{n+1} - T_n$ are used; the sample-path-based approach used here is easier to implement on-line (e.g., see the definition in (2.84)).

The discounted reward with a discount factor $\beta > 0$ for semi-Markov processes is defined as

$$\eta_\beta(i) = \lim_{T \to \infty} E \left[\int_0^T \beta \exp(-\beta t) f(X_t, Y_t) dt \middle| X_0 = i \right], \qquad T_0 = 0. \quad (2.93)$$

Similar to the discrete case in (2.31), the weighting factor in (2.93) is also normalized: $\int_0^\infty \beta \exp(-\beta t) dt = 1$. The performance potential for the discounted reward criterion is

$$g_\beta(i) = \lim_{T \to \infty} E \left\{ \int_0^T \exp(-\beta t)[f(X_t, Y_t) - \eta] dt \middle| X_0 = i \right\}, \qquad i \in \mathcal{S}.$$

The sensitivity analysis of the discounted reward for semi-Markov processes involves an *equivalent Markov process*. We refer readers to [57] for technical details.

2.4 Perturbation Analysis of Queueing Systems

The early works on perturbation analysis (PA) focused on queueing systems. The idea of PA was first proposed in [144] for the buffer allocation problem in a serial production line and was first studied for queueing networks in [141]. The special structure of queueing systems, especially the interactions among different customers or different servers, makes PA a very efficient tool for estimating the performance derivatives with respect to the mean service times based on a single sample path. This section contains an overview of the main results of PA of queueing systems.

The main difference between PA of Markov chains and PA of queueing systems is that in the former, a perturbation is a "jump" on a sample path from one state to another due to parameter changes, while, in the latter, it is a small (infinitesimal) delay in a customer's transition time. Some queueing (such as the Jackson-type) networks can be modelled by Markov processes and therefore the theory and algorithms developed for Markov processes can be applied. However, because of the special features of a queueing system, the performance derivatives with respect to service time changes can be obtained by a much more efficient and more intuitive approach, which applies to non-Markov queueing systems as well.

The dynamic nature of a system's behavior is explored more clearly in PA of queueing systems. Its basic principle can be described as follows: a small increase in the mean service time of a server *generates* a series of small delays, called *perturbations*, in the service completion times of the customers served by that server. Each such perturbation of a customer's service completion time will cause delays in the service completion times of other customers (at the same server or at other servers). In other words, a perturbation will be *propagated* through the system due to the interactions among customers and servers. Thus, a perturbation will affect the system performance through propagation. The average effect of a perturbation on the system performance can be measured by a quantity called the *perturbation realization factor (PRF)*. Finally, the effect of a change in the mean service time of a server equals the sum of the effects of all the perturbations generated on the service completion times of the server due to this change in its mean service time. The above description is precisely captured by *three fundamental rules of PA*:

1. Perturbation generation;
2. Perturbation propagation;
3. Perturbation realization.

These rules will be discussed in subsequent subsections. In PA of Markov chains, the perturbations (jumps) are generated according to (2.2) and (2.4); the perturbation realization is illustrated by Figure 2.6 and measured by (2.5).

However, as we will see, the "propagation" effect in Markov chains is not as explicit as in queueing networks.

Problem Description

Consider a closed Jackson network (cf. Appendix C.2) with M servers and N customers. The service times of every server in the network are independently and exponentially distributed. Let $\bar{s}_i$ be the mean service time of server i, $i = 1, \dots, M$, and let $q_{i,j}$, $i, j = 1, \dots, M$, be the routing probabilities. $Q = [q_{i,j}]$ is the routing probability matrix. The system state can be denoted as $\boldsymbol{n} = (n_1, n_2, \dots, n_M)$, where n_i is the number of customers in server i. For a closed network with M servers and N customers, we have $\sum_{i=1}^{M} n_i = N$. The state space is $\mathcal{S} = \{\text{all } \boldsymbol{n} : \ \sum_{i=1}^{M} n_i = N\}$. The system state at time t is denoted as $\boldsymbol{N}(t) = (n_1(t), \dots, n_M(t))$. The system can be modelled by a Markov process $\boldsymbol{X} = \{\boldsymbol{N}(t), t \geq 0\}$. Let T_l, $l = 0, 1, \dots$, be the lth transition time of $\boldsymbol{X}$, counting the customer transitions at all the servers. Figure 2.12 illustrates a sample path of a three-server five-customer closed queueing network. The vertical dashed arrows signal the customer transitions among servers, and each of the three staircase-like curves indicates the evolution of a server in the network.

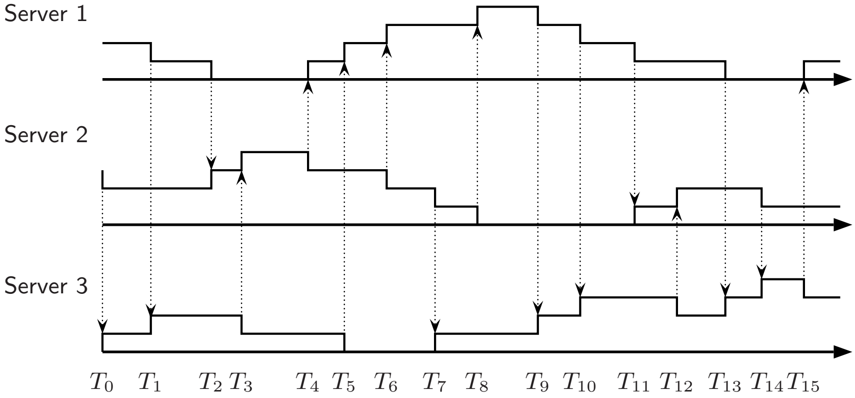


Fig. 2.12. A Sample Path of a Closed Queueing Network with $M = 3$ and $N = 5$

Let $f : \mathcal{S} \to \mathcal{R}$ be a reward (or cost) function. The system performance is defined as the long-run average reward

$$\eta^{(f)} = \lim_{L \to \infty} \frac{1}{L} \int_0^{T_L} f[\boldsymbol{N}(t)] dt, \qquad \text{w.p.1}, \tag{2.94}$$

where T_L is the Lth transition time of the system. In this section, we use the superscript "(f)" to explicitly denote the dependency of a quantity on f for clarity. For closed Jackson networks in which a customer can reach any server in the network while circulating in the network (irreducible networks), the state process $\boldsymbol{N}(t)$ is an ergodic Markov process, and the limit in (2.94) exists with probability 1 and does not depend on the initial state. Set

$$F_L = \int_0^{T_L} f[\boldsymbol{N}(t)]dt.$$

Then, we have

$$\eta^{(f)} = \lim_{L\to\infty} \frac{F_L}{L}.$$

With L being the number of customers' service completions in the period of $[0, T_L]$, the performance measure defined in (2.94) is the *customer average*. These types of performance measures cover a wide range of applications. For example, if $f(\boldsymbol{n}) = I(\boldsymbol{n}) \equiv 1$ for all $\boldsymbol{n} \in \mathcal{S}$, then $F_L = T_L$ and

$$\eta^{(I)} = \lim_{L\to\infty} \frac{T_L}{L} = \frac{1}{\eta}, \tag{2.95}$$

where $\eta = \lim_{L\to\infty} \frac{L}{T_L}$ is the system throughput (the number of service completions per unit of time). If $f(\boldsymbol{n}) = n_i$, then F_L is the area underneath the sample path of server i. Let L_i be the number of service completions at server i in $[0, T_L]$. Then,

$$\eta^{(f)} = \lim_{L\to\infty} \frac{F_L}{L} = \left(\lim_{L\to\infty} \frac{L_i}{L}\right)\left(\lim_{L\to\infty} \frac{F_L}{L_i}\right) = v_i \bar{\tau}_i,$$

where v_i is the visit ratio of server i (see (C.5) in Appendix C), satisfying $v_i = \sum_{j=1}^M v_j q_{j,i}$ and normalized to $\sum_{k=1}^M v_i = 1$, and $\bar{\tau}_i$ is the mean response time (waiting time + service time) of a customer at server i. Similarly, we have

$$\eta^{(f)} = \lim_{L\to\infty} \frac{F_L}{L} = \left(\lim_{L\to\infty} \frac{T_L}{L}\right)\left(\lim_{L\to\infty} \frac{F_L}{T_L}\right) = \eta^{(I)} \bar{n}_i,$$

where $\bar{n}_i$ is the average number of customers at server i.

Another type of performance measure is the long-run time-average reward defined as

$$\eta_T^{(f)} = \lim_{L\to\infty} \frac{1}{T_L} \int_0^{T_L} f[\boldsymbol{N}(t)]dt,$$

which can be easily converted to customer averages as follows:

$$\eta_T^{(f)} = \eta\eta^{(f)} = \frac{\eta^{(f)}}{\eta^{(I)}}.$$

Now suppose that the mean service time of one of the servers, say server v, changes from $\bar{s}_v$ to $\bar{s}_v + \Delta\bar{s}_v$. We call the closed network with $\bar{s}_i$, $i = 1, 2, \ldots, M$, the original network, and the network with the changed mean service time $\bar{s}_v + \Delta\bar{s}_v$ and $\bar{s}_i$, $i \neq v$, the perturbed network. A sample path of the original network is called an *original sample path*, and a sample path of the perturbed network is called a *perturbed sample path.*

Given a sample path of a network, its average reward $\eta^{(f)}$ can be easily estimated by simple calculation. The goal of PA is to obtain an estimate for the performance derivatives $\frac{d\eta^{(f)}}{d\bar{s}_v}$, $v = 1, 2, \ldots, M$, by observing and analyzing an original sample path. This is shown in Figure 2.13, in which we use θ to denote a generic parameter.

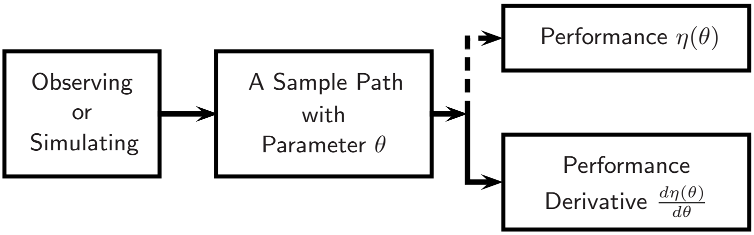

Observing or Simulating
A Sample Path with Parameter θ
Performance $\eta(\theta)$
Performance Derivative $\frac{d\eta(\theta)}{d\theta}$

Fig. 2.13. The Goal of Perturbation Analysis

2.4.1 Constructing a Perturbed Sample Path

As in Markov chains, the first step in PA of queueing systems is to construct a perturbed sample path by using an original one.

Suppose that we are given an original sample path with transition times T_l, $l = 0, 1, \ldots$. Let T'_l be the lth transition time on the corresponding perturbed path, $l = 0, 1, \ldots$. Suppose that the lth transition time is a service completion time of server i. Then, $\Delta T_l := T'_l - T_l$ is called the *perturbation of server* i at time T_l; it is also called the perturbation of the customer that completes the service at server i at T_l.

Perturbation Generation

First, we study how the change in the mean service time of a server affects every customer's service time at that server. In general, let $\bar{s}$ be the mean service time of a server with an exponentially distributed service time. Then, the service time of a customer at that server, denoted as s, has the following distribution:

$$\Phi(s) = 1 - \exp\left(-\frac{s}{\bar{s}}\right).$$

In simulation, we use the inverse-transform method to generate the service times (shown in Figure A.2 and reproduced in Figure 2.14). First, we generate a uniformly distributed random number ξ in $[0, 1)$. Then, we set

$$s = \Phi^{-1}(\xi) = -\bar{s}\ \ln(1-\xi). \tag{2.96}$$

It is well known that s in (2.96) is exponentially distributed with mean $\bar{s}$. Assume that the mean service time changes to $\bar{s}+\Delta\bar{s}$ (for the sake of discussion, we may assume that $\Delta\bar{s} > 0$). Then, with the same random variable ξ, the service time in (2.96) changes to

$$s + \Delta s = -(\bar{s}+\Delta\bar{s})\ \ln(1-\xi).$$

Thus, we have

$$\Delta s = -\Delta\bar{s}\ln(1-\xi) = \frac{\Delta\bar{s}}{\bar{s}}s = \kappa s, \qquad \kappa := \frac{\Delta\bar{s}}{\bar{s}}. \tag{2.97}$$

That is, the service time of every customer at the perturbed server will increase by an amount Δs (> 0) shown in (2.97); in other words, the service completion time of every customer at the server will be delayed by Δs (> 0). We call (2.97) the *perturbation generation rule* [142]:

The Perturbation Generation Rule:

At the perturbed server, because of the change in the mean service time $\Delta\bar{s}$, every customer's service completion time obtains a perturbation of Δs, shown in (2.97), on the sample path.

This perturbation obtained during a customer's service period is in addition to the perturbation(s) previously obtained by the server before the customer starts its service.

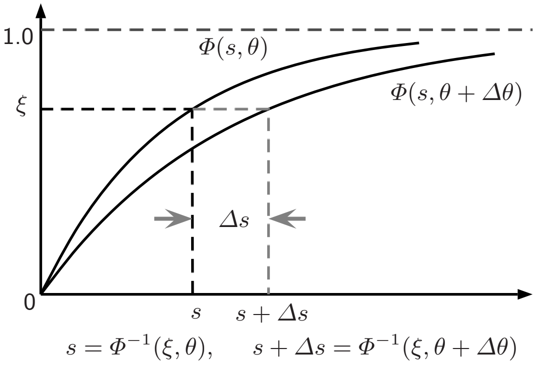


Fig. 2.14. The Perturbation Generation Rule

The inverse-transform method can be used to derive the perturbation generation rule for other service distributions. Let $\Phi(s, \theta)$ be the distribution function of the service times of the customers at a server, which depends on a parameter θ. With the inverse-transform method, we determine a customer's service time by using the inverse function of the distribution function:

$$s = \Phi^{-1}(\xi, \theta) = \sup\{s : \ \Phi(s, \theta) \leq \xi\},$$

where ξ is a uniformly distributed random variable on $[0, 1)$. Suppose that the distribution parameter θ changes to $\theta + \Delta\theta$. Then, the service time of the customer changes to

$$s + \Delta s = \Phi^{-1}(\xi, \theta + \Delta\theta).$$

We have

$$\begin{aligned} \Delta s &= \Phi^{-1}(\xi, \theta + \Delta\theta) - \Phi^{-1}(\xi, \theta) \\ &\approx \left. \frac{\partial \Phi^{-1}(\xi, \theta)}{\partial \theta} \right|_{\xi = \Phi(s,\theta)} \Delta\theta = \left. \frac{\partial s}{\partial \theta} \right|_{\xi = \Phi(s,\theta)} \Delta\theta. \end{aligned} \tag{2.98}$$

Δs is the *perturbation generated* during the service period because of $\Delta\theta$. The same random variable ξ is used for both s and $s + \Delta s$. Pictorially, the perturbation generation rule is illustrated in Figure 2.14.

In practice, calculating the partial derivative $\frac{\partial \Phi^{-1}(\xi,\theta)}{\partial \theta}$ may require a relatively large amount of computation. However, in most applications, such as in communication systems, the packet length distribution, $length = \Phi^{-1}(\xi)$, is fixed, and one can only change the transition rate μ. The service (transition) time is $s = \frac{length}{\mu} = \frac{1}{\mu}\Phi^{-1}(\xi)$. Therefore, for service rate μ, we have

$$\begin{aligned} \Delta s &\approx -\frac{\Delta\mu}{\mu^2}\Phi^{-1}(\xi) = -\frac{\Delta\mu}{\mu}s \\ &= \kappa s, \qquad \kappa = -\frac{\Delta\mu}{\mu}, \end{aligned} \tag{2.99}$$

which is in the same form as (2.97) for the mean service time of the exponential distribution.

Perturbation Propagation

A perturbation of one customer, or one server, will affect the transition times of other customers, or other servers, in the network. Figure 2.15 illustrates the interaction between two servers. Suppose that the first customer in server 1 obtains a perturbation Δ at time T_1; i.e., its service completion time is delayed by Δ. Apparently, the service starting time of the next customer at the same server will be delayed by Δ and its service completion time will also be delayed by the same amount Δ at T_2. In addition, at T_1, because server 2 was idle and was waiting for a customer arriving from server 1, the service starting time of server 2 at T_1 and its completion time at T_3 will also be delayed by Δ. We summarize the above discussion in two *perturbation propagation rules* [142]:

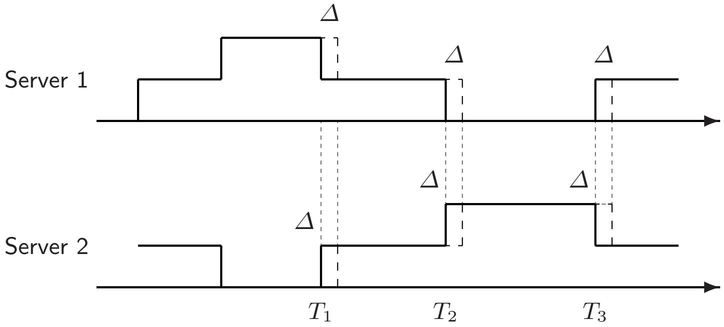


Fig. 2.15. Perturbation Propagation

The Perturbation Propagation Rules:

i. A server keeps its perturbation until it meets an idle period (or the perturbation of a customer's service completion time is propagated to the next customer in the server until the server meets an idle period).

ii. The perturbation of one server will be propagated to another server if a customer at the former moves to the latter and terminates an idle period of the latter server.

The first rule implies that when a server meets an idle period, the server's original perturbation is lost. The second rule implies that after the idle period, the server will acquire a perturbation propagated from another server. That is, after an idle period, a server's perturbation always equals that of the server that terminates the idle period. A special case is illustrated in Figure 2.16, in which server 1 has a perturbation $\varDelta_1 = \varDelta$ at T_1, but after the idle period, at T_2, the server acquires the perturbation from server 2, which is $\varDelta_2 = 0$. Thus, the perturbation $\varDelta_1$ of server 1 at T_1 is lost after the idle period at T_2. This explains how a non-perturbed server can be viewed as a server having a perturbation 0 in perturbation propagation.

Note that we assume that the perturbation can be as small as we wish (*infinitesimal perturbation*). Thus, we can always assume that the perturbation is smaller than the length of the idle period. See Section 2.4.4 for more details.

Constructing a Perturbed Path

Now we return to the closed Jackson network with M servers and N customers. Suppose that we are given a sample path of the original system with

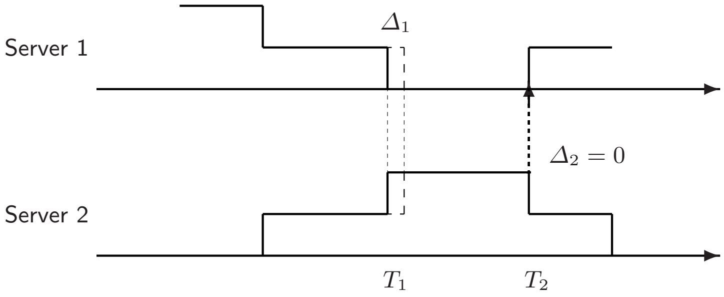


Fig. 2.16. Perturbation Propagation for $\Delta = 0$

mean service times $\bar{s}_i$, $i = 1, 2, \ldots, M$, and one server's (server v) mean service time is perturbed from $\bar{s}_v$ to $\bar{s}_v + \Delta\bar{s}_v$.

From the perturbation generation and propagation rules, we can efficiently determine the perturbations of all the servers and therefore construct a perturbed sample path on an original one without simulating the perturbed system again. We may simply generate a perturbation on the original sample path according to (2.97) whenever a customer completes its service at server v, and then propagate it along the original sample path according to the two propagation rules. Note that we can propagate all the perturbations at a server altogether. This leads to the following simple algorithm for determining the perturbations of all servers on the sample path at any time:

Algorithm 2.1. (Constructing a Perturbed Sample Path)

Given an original sample path for a closed Jackson network:

i. Initialization: Set $\Delta_i := 0$, $i = 1, 2, \ldots, M$;
ii. (Perturbation generation) At the kth service completion time of server v, set $\Delta_v := \Delta_v + s_{v,k}$, $k = 1, 2, \ldots$, $s_{v,k}$ is the service time of the customer;
iii. (Perturbation propagation) After a customer from server i terminates an idle period of server j, set $\Delta_j := \Delta_i$, $i, j = 1, 2, \ldots, M$.

The perturbation of server i is $\kappa\Delta_i$, $i = 1, 2, \ldots, M$.

In the algorithm, Δ_i denotes the (accumulated) perturbation of server i, $i = 1, 2, \ldots, M$. The perturbation of every server is updated whenever it starts a new busy period, and, in addition, the perturbation of the perturbed server is also updated whenever it completes its service to a customer. Because all the perturbations generated and propagated are proportional to $\kappa = \frac{\Delta\bar{s}_v}{\bar{s}_v}$, at any time the perturbation at any server in the network must be proportional to κ. Therefore, for simplicity, in the algorithm, we use $s_{v,k}$ instead of $\kappa s_{v,k}$ as the perturbation generated. Thus, the exact perturbation corresponding to $\Delta \bar{s}_v$ at any server i should be $\kappa \Delta_i$, $i = 1, \ldots, M$. The algorithm determines the perturbations of all the transition times of all servers (i.e., T_l, $l = 1, 2, \ldots$) at the perturbed path. The transition times of the perturbed path equal those of the original path plus the perturbation of the corresponding server.

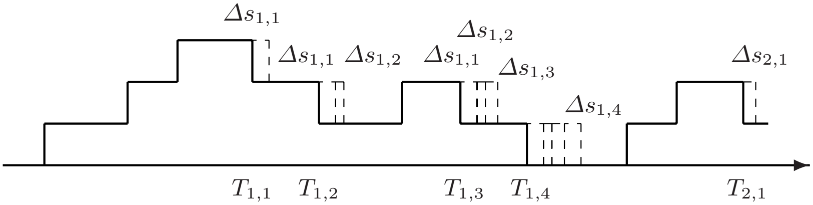


Fig. 2.17. A Perturbed Sample Path for an M/G/1 Queue

Example 2.4. To illustrate perturbation propagation within the same server, we consider a single server queue, in which there is no perturbation propagation among different servers. In such a system, the third step in Algorithm 2.1 is not implemented, and the perturbation of the server is reset to zero at the beginning of every new busy period. Actually, a single server queue is an open network, and the arriving customers can be viewed as from a source that is never perturbed. A sample path of such a single server queue (may be viewed as an M/M/1 or an M/G/1 queue) and its corresponding perturbed path constructed by Algorithm 2.1 are shown in Figure 2.17.

The figure illustrates the first busy period of the sample path, in which there are four customers served by the server. In the kth busy period, the ith customer's service time is denoted as $s_{k,i}$, and its departure time is denoted as $T_{k,i}$, $k, i = 1, 2, \ldots$. At the first customer's departure time $T_{1,1}$, a perturbation $\Delta s_{1,1} = \kappa s_{1,1}$ is generated according to (2.98) or (2.99). This perturbation is propagated to the departure times of the subsequent customers in the same busy period, $T_{1,2}$, $T_{1,3}$, and $T_{1,4}$. At $T_{1,2}$, another perturbation $\Delta s_{1,2} = \kappa s_{1,2}$ is generated; thus, the total perturbation at $T_{1,2}$ is $\Delta s_{1,1} + \Delta s_{1,2}$. This perturbation propagates to $T_{1,3}$ and $T_{1,4}$; and so on. In general, the perturbation of the ith departure time in the kth busy period is

$$\Delta T_{k,i} = \sum_{l=1}^{i} \Delta s_{k,l}, \tag{2.100}$$

where $\Delta s_{k,l} = \kappa s_{k,l}$ is the perturbation of the lth customer's service time in the kth busy period, generated according to (2.98) or (2.99). □

Figure 2.17 also illustrates a fundamental fact: The simple rules for perturbation propagation hold only if the perturbation accumulated at the end of a busy period is smaller than the length of the idle period following the busy period. For the time being, we may think that we can always choose $\Delta\bar{s}$ or $\Delta\theta$ small enough such that this condition holds. For a rigorous discussion, see Section 2.4.4.

Example 2.5. Suppose that we are given an original sample path of a three-server five-customer closed network shown in Figure 2.12, and server 2's mean service time is perturbed from $\bar{s}_2$ to $\bar{s}_2 + \Delta\bar{s}_2$. We may construct a perturbed sample path by following the perturbation generation and propagation rules, as shown in Figure 2.18. The top figure shows the original path plus the perturbations at all transition instants; and the bottom figure shows the perturbed path thus constructed, in which $T'_l = T_l + \Delta T_l$, with ΔT_l being the perturbation of the transition instant T_l, $l = 0, 1, \ldots, 15$.

There are five perturbations generated, denoted as perturbations Δs_1, Δs_2, Δs_3, Δs_4, and Δs_5 (for simplicity, we omitted the subscript denoting server 2, e.g., we write $\Delta s_{2,1} = \Delta s_1$, etc.) and differentiated by different grays shown in the figure. The five perturbations are induced during the first five customers' service times at the perturbed server, server 2. They are generated according to the perturbation generation rule (2.98).

As shown in the figure, Perturbation Δs_1 obtained at T_4 by server 2 is propagated to server 1 immediately since the customer at server 2 terminates an idle period of server 1 at T_4. This perturbation is also propagated to the subsequent service completion times of server 2, T_6, T_7 and T_8. At T_6, server 2 obtains another perturbation Δs_2 for its second customer, resulting in a total perturbation of $\Delta T_6 = \Delta s_1 + \Delta s_2$. Similarly, we have $\Delta T_7 = \Delta s_1 + \Delta s_2 + \Delta s_3$ and $\Delta T_8 = \Delta s_1 + \Delta s_2 + \Delta s_3 + \Delta s_4$. As shown in the figure, ΔT_7 is propagated to server 3 through an idle period. The perturbation that is propagated to server 1 at T_4, Δs_1, is also propagated to the subsequent customers' service completion times, T_9, T_{10}, T_{11}, and T_{13}, in the same busy period of server 1. Likewise, the perturbation propagated to server 3 at T_7, $\Delta s_1 + \Delta s_2 + \Delta s_3$, is also propagated to the subsequent customers' service completion times, T_{12} and T_{15}, in the same busy period of server 3.

The perturbation that server 2 acquired in the first busy period $\Delta T_8 = \Delta s_1 + \Delta s_2 + \Delta s_3 + \Delta s_4$ is lost after the idle period starting from T_8. Indeed, at the beginning of the next busy period T_{11}, server 2 acquires a perturbation $\Delta T_{11} = \Delta s_1$ through propagation from server 1. There is another perturbation, Δs_5, generated during the service time of the first customer in the second busy period of server 2, resulting in a total perturbation of $\Delta T_{14} = \Delta s_1 + \Delta s_5$ for server 2 at T_{14}. Note that although the arrival time to server 1 at T_8 is delayed by $\Delta s_1 + \Delta s_2 + \Delta s_3 + \Delta s_4$, its effect is temporary: it does not affect any other service completion time at server 1 at all. The same statement holds for the delays in other arrival times except for those arrivals that start a new busy period.

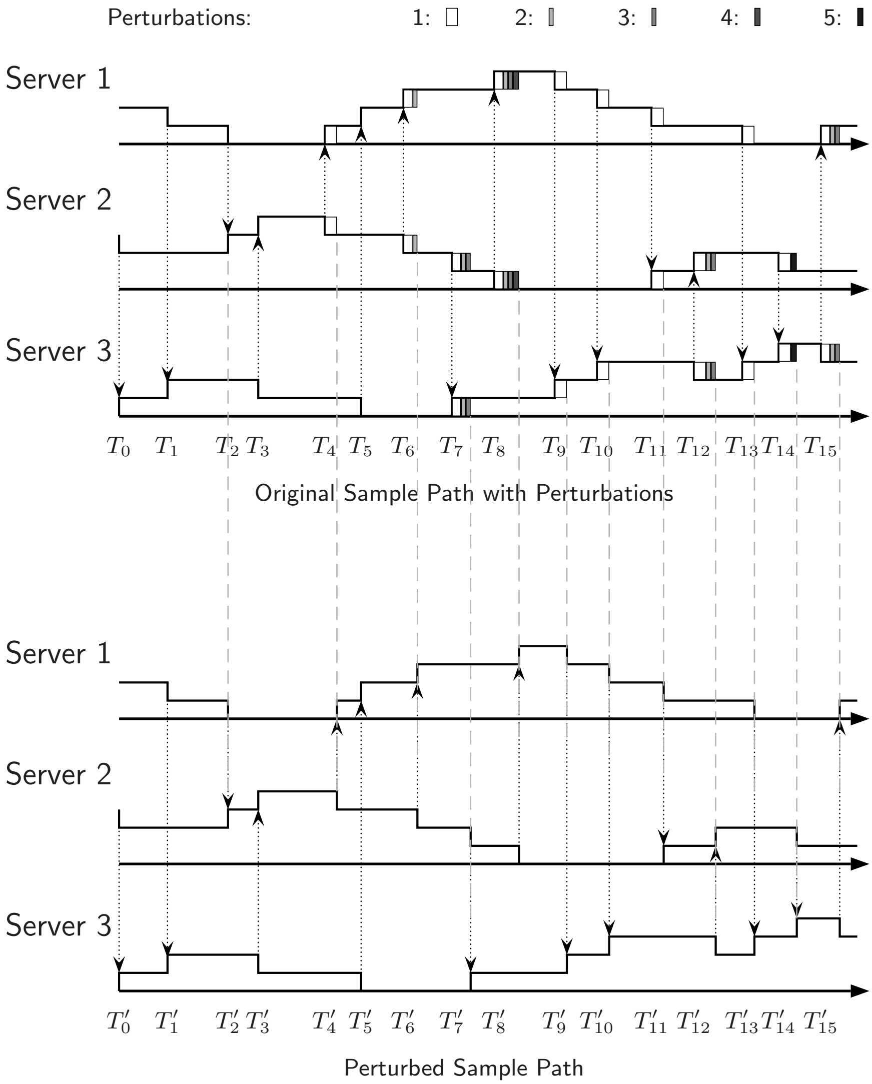

Perturbations:
1: □
2: ▯
3: ▯
4: ▮
5: ▮
Server 1
Server 2
Server 3
T_0 T_1 T_2 T_3 T_4 T_5 T_6 T_7 T_8 T_9 T_{10} T_{11} T_{12} T_{13} T_{14} T_{15}
Original Sample Path with Perturbations
Server 1
Server 2
Server 3
T'_0 T'_1 T'_2 T'_3 T'_4 T'_5 T'_6 T'_7 T'_8 T'_9 T'_{10} T'_{11} T'_{12} T'_{13} T'_{14} T'_{15}
Perturbed Sample Path

Fig. 2.18. A Perturbed Sample Path of the Network in Figure 2.12

Again, the perturbation propagated to server 1 in the second busy period, Δs_1, is lost after the idle period starting from T_{13}. After that, server 1 acquires a perturbation $\Delta s_1 + \Delta s_2 + \Delta s_3$ from server 3 through propagation at T_{15}.

It is interesting to note that starting from T_7, every server acquires the perturbation Δs_1. We say that Δs_1 is realized at T_7 by the network. In contrast, starting from T_9, no server has the perturbation Δs_4. We say that Δs_4 is lost by the network at T_9 (see Section 2.4.2). □

Calculating the Performance Derivatives

With the perturbed sample path constructed by Algorithm 2.1, the performance of the perturbed system can be calculated. As an example, we consider the system throughput. Recall that T_L is the Lth transition time of a queueing system. Assume that $L >> 1$. Then, the overall system throughput (the number of customers served by all the servers in the network per unit of time) is defined as

$$\eta = \lim_{L \to \infty} \frac{L}{T_L} \approx \frac{L}{T_L}.$$

In the perturbed system with $\bar{s}_v$ changed to $\bar{s}_v + \Delta \bar{s}_v$, it takes $T_L + \Delta T_L$ to finish the L transitions, with $\Delta T_L = \kappa \Delta_u$, $\kappa = \frac{\Delta \bar{s}_v}{\bar{s}_v}$, where u denotes the server for which T_L is the service completion time, and Δ_u is its perturbation at T_L determined by Algorithm 2.1 (in which κ is set to be one). The throughput of the perturbed system is

$$\eta + \Delta\eta \approx \frac{L}{T_L + \Delta T_L} \approx \frac{L}{T_L}(1 - \frac{\Delta T_L}{T_L}) = \eta(1 - \frac{\Delta T_L}{T_L}).$$

Thus, we have

$$\Delta\eta \approx -\eta \frac{\Delta T_L}{T_L},$$

and

$$\frac{\bar{s}_v}{\eta} \frac{\Delta\eta}{\Delta \bar{s}_v} \approx -\frac{\bar{s}_v}{\Delta \bar{s}_v} \frac{\Delta T_L}{T_L} = -\frac{\Delta_u}{T_L}.$$

Therefore, the *elasticity* (or the *normalized derivative*) of η with respect to $\bar{s}_v$ can be estimated on a sample path with PA as follows.

$$\frac{\bar{s}_v}{\eta} \frac{\partial\eta}{\partial \bar{s}_v} \approx -\frac{\Delta_u}{T_L}, \tag{2.101}$$

which does not depend on κ!

To obtain the derivatives of the throughput, in addition to the throughput itself, the algorithm adds only three clauses to the simulation program, one for perturbation generation, one for perturbation propagation (see Algorithm 2.1), and one for calculating the normalized derivative according to (2.101); and it adds only about 5% of computation time [141]. The following example illustrates the accuracy of this algorithm.

Example 2.6. Consider a closed Jackson network with $M = 6$, $N = 12$; the mean service times of the servers are 30, 40, 50, 55, 45, and 35, respectively; and the routing probability matrix is

$$Q = \begin{bmatrix} 0.00 & 0.10 & 0.20 & 0.15 & 0.35 & 0.20 \\ 0.25 & 0.00 & 0.15 & 0.10 & 0.10 & 0.40 \\ 0.35 & 0.15 & 0.00 & 0.25 & 0.25 & 0.00 \\ 0.25 & 0.25 & 0.10 & 0.00 & 0.20 & 0.20 \\ 0.00 & 0.20 & 0.25 & 0.15 & 0.00 & 0.40 \\ 0.40 & 0.30 & 0.00 & 0.15 & 0.15 & 0.00 \end{bmatrix}.$$

We ran a simulation for $L = 500,000$ transitions and applied the PA Algorithm 2.1 to the simulation. The resulting elasticities of the system throughput with respect to each mean service time given by (2.101) and the theoretical values of these elasticities (calculated by (C.19) and (C.15)) are shown in Table 2.8. □

$-\frac{\bar{s}_i}{\eta}\frac{\partial\eta}{\partial\bar{s}_i}$	$i = 1$	2	3	4	5	6
PA estimate	0.0906	0.1374	0.1025	0.2131	0.2736	0.1828
Theoretical	0.0915	0.1403	0.0980	0.2087	0.2812	0.1802

Table 2.8. Elasticities in Example 2.6

Now, we consider the average reward defined with any general reward function f in (2.94):

$$\eta^{(f)} = \lim_{L\to\infty} \frac{1}{L} \int_0^{T_L} f[\boldsymbol{N}(t)]dt = \lim_{L\to\infty} \frac{F_L}{L}, \tag{2.102}$$

where $\boldsymbol{N}(t)$ denotes the state process and

$$F_L = \int_0^{T_L} f[\boldsymbol{N}(t)]dt.$$

The computation of the performance derivative $\frac{\partial\eta^{(f)}}{\partial\bar{s}_i}$ involves more than that of the derivative of the system throughput $\frac{\partial\eta}{\partial\bar{s}_i}$. It depends not only on the final perturbation $\varDelta_u$, as shown in (2.101), but also on the perturbations of every transition time. We need to modify Algorithm 2.1 as follows:

Algorithm 2.2. (Calculating the Performance Derivatives)

Given an original sample path for a closed Jackson network:

i. Initialization: Set $\Delta_i := 0$, $i = 1, 2, \dots, M$, and $\Delta F := 0$;

ii. (Perturbation Generation and Propagation) Same as steps ii and iii in Algorithm 2.1, which determine the perturbations of T_l, ΔT_l, $l = 1, 2, \dots;$

iii. (Update ΔF) At every transition time T_l, $l = 1, 2 \dots$, set $\Delta F := \Delta F + [f(\boldsymbol{n}) - f(\boldsymbol{n}')]\Delta T_l$, where $\boldsymbol{n} = \boldsymbol{N}(T_{l-})$ and $\boldsymbol{n}' = \boldsymbol{N}(T_l)$ are the system states before and after the transition, respectively.

Similar to Algorithm 2.1, κ is also set to be one in Algorithm 2.2. Let ΔF_L be the perturbation obtained by the algorithm at T_L. Then, the real perturbation of F_L for the system is $\kappa \Delta F_L$, with $\kappa = \frac{\Delta \bar{s}_v}{\bar{s}_v}$. Thus, when L is sufficiently large, from (2.102), we have $\Delta \eta^{(f)} = \kappa \frac{\Delta F_L}{L}$. From this, we obtain

$$\frac{\bar{s}_v}{\eta^{(I)}} \frac{\partial \eta^{(f)}}{\partial \bar{s}_v} \approx \frac{\bar{s}_v}{\eta^{(I)}} \frac{\Delta \eta^{(f)}}{\Delta \bar{s}_v} = \frac{\Delta F_L}{T_L}.$$

Finally, both Algorithms 2.1 and 2.2 can be implemented on line; i.e., there is no need to store the history of the sample path. Ref. [64] contains some simulation examples for Algorithm 2.2, applied to mean response times.

2.4.2 Perturbation Realization

We derived the PA algorithms for performance derivatives in the previous subsection. In this subsection, we start a more rigorous study of PA.

We first introduce the fundamental concept in PA: the *perturbation realization*. We show that, on average, the final effect of a single perturbation on the system performance (more precisely, on F_L, $L >> 1$ in (2.102)) can be measured by a quantity called the *perturbation realization factor*. Therefore, roughly speaking, the effect of a change in a system's parameter on the performance equals the sum of the realization factors of all the perturbations that are induced by the parameter change. This general principle is the same as in PA of Markov chains. The difference is that a perturbation for a queueing system is a small (infinitesimal) delay in time and that for a Markov chain is a state "jump". Historically, however, this principle was first proposed for PA of queueing systems [45, 49, 50, 51, 113, 141], and was extended later to Markov systems [62, 70].

Perturbation Realization

Consider the M-server closed Jackson network discussed in Section 2.4.1. The performance is defined as (2.94). To study the effect of a single perturbation

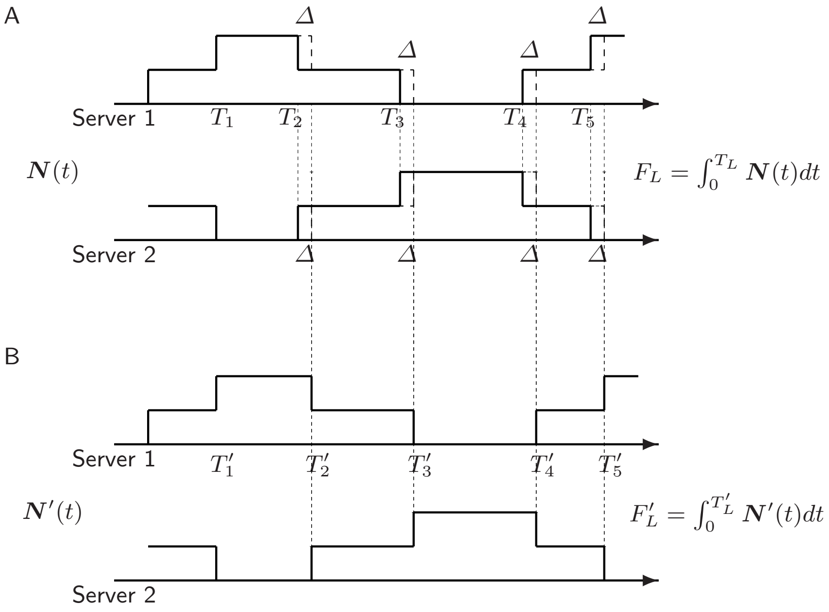


Fig. 2.19. A Sample Path and its Perturbed Counterpart

on the performance $\eta^{(f)}$, we assume that at some time, a perturbation Δ is generated at a server (e.g., in Figure 2.15 a perturbation Δ is generated at server 1 at T_1). As explained in Section 2.4.1, this perturbation will be propagated along a sample path. To study the effect of this single perturbation, we assume that there is no other perturbation generated on the sample path. During propagation, some servers in the network acquire this perturbation (e.g., in Figure 2.15, server 2 obtains a perturbation at T_1); others may lose the perturbation obtained before (e.g., in Figure 2.16, server 1 loses its perturbation at T_2). During propagation, every server has either perturbation Δ or perturbation 0 (no perturbation).

If, through propagation, every server in the network acquires the perturbation Δ, we say that the perturbation is *realized* by the network. After the perturbation is realized, the perturbed sample path is the same as the original one except that the entire sample path is shifted to the right by the amount of Δ. That is, there is an L^*, such that $T'_l = T_l + \Delta$ for all $l \geq L^*$. If, through propagation, every server in the network loses its perturbation (or acquires a perturbation of 0), we say that the perturbation is *lost* by the network. After the perturbation is lost, the perturbed sample path is exactly the same as the original one. That is, there is an L^*, such that $T'_l = T_l$ for all $l \geq L^*$. Apparently, whether a perturbation is realized or lost is random and depends on the sample path.

The solid lines in Figure 2.19.A illustrate a sample path $\boldsymbol{N}(t)$ of a two-server two-customer cyclic queueing network consisting of transition instants T_1 to T_5. A perturbation Δ is generated at server 1 at T_2, which is propagated to server 2 at T_2, and after T_2 all the servers have the same perturbation Δ, and the perturbation is realized by the network. The perturbed sample path corresponding to this perturbation is shown in Figure 2.19.B.

A closed queueing network is called *irreducible* if a customer at any server may visit any other server in the network, either directly, or by going through other servers. That is, for any pair of $i, j \in \{1, 2, \ldots, M\}$, there exists a sequence of integers, $k_1, k_2, \ldots, k_m \in \{1, 2, \ldots, M\}$, such that $q_{i,k_1} q_{k_1,k_2} \cdots q_{k_m,j} > 0$. Such a routing probability matrix Q is also called *irreducible*. The following theorem indicates that a closed irreducible network will eventually "settle down" after being perturbed by a small perturbation.

Theorem 2.1. A perturbation in an irreducible closed Jackson network will either be realized or lost by the network with probability 1.

Proof. Since the network is irreducible, the state process $\boldsymbol{N}(t)$ will visit any state. In particular, with probability 1 every sample path will eventually visit state $(N, 0, \ldots, 0)$; i.e, all customers are at server 1. If at that time server 1 has the perturbation, then after that time, all the servers will have the same perturbation; i.e., the perturbation is realized. On the other hand, if at that time server 1 has no perturbation, then after it all the servers in the network will have no perturbation; i.e., the perturbation is lost. □

The probability that a perturbation is realized is called the *perturbation realization probability*. It depends on the system state. The realization probability of a perturbation of server i when the system is in state $\boldsymbol{n}$ is denoted as $c(\boldsymbol{n}, i)$, $\boldsymbol{n} \in \mathcal{S}$, $i = 1, 2, \ldots, M$.

Example 2.7. In Figure 2.18, the perturbation generated at T_4, Δs_1, is realized by the network at T_7. The perturbation generated at T_8, Δs_4, is lost at T_{11}. The other three perturbations, Δs_2, Δs_3, and Δs_5, have not been either realized or lost at T_{15}. Whether they will be realized or lost depends on the future evolution of the sample path. □

Perturbation Realization Factors

The effect of a perturbation on the long-run average reward $\eta^{(f)}$ defined in (2.94) can be studied by using the concept of perturbation realization. We first define the *realization factor* of a perturbation Δ of server i in state $\boldsymbol{n}$ for $\eta^{(f)}$ as (cf. (2.6) for realization factors for Markov chains):

$$c^{(f)}(\boldsymbol{n}, i) = \lim_{L\to\infty} E\left(\frac{\Delta F_L}{\Delta}\right) = \lim_{L\to\infty} E\left(\frac{F'_L - F_L}{\Delta}\right)$$
$$= \lim_{L\to\infty} E\left\{\frac{1}{\Delta}\left\{\int_0^{T'_L} f[\boldsymbol{N}'(t)]dt - \int_0^{T_L} f[\boldsymbol{N}(t)]dt\right\}\right\}, \quad (2.103)$$

where F'_L is measured on the perturbed path generated by the propagation of this perturbation Δ (see Figure 2.19). It is clear that the realization factor $c^{(f)}(\boldsymbol{n}, i)$ measures the average effect of a perturbation at $(\boldsymbol{n}, i)$ on F_L in (2.94) as $L \to \infty$.

Recall that if a perturbation is realized, then there is an integer L^*, such that $T'_L = T_L + \Delta$ for all $L \geq L^*$, and if a perturbation is lost, then there is an L^*, such that $T'_L = T_L$ for all $L \geq L^*$. In both cases, there is an L^* (depending on the sample path) such that

$$\int_{T_{L^*}}^{T_L} f[\boldsymbol{N}(t)]dt - \int_{T'_{L^*}}^{T'_L} f[\boldsymbol{N}'(t)]dt = 0,$$

for all $L \geq L^*$ (in Figure 2.19, $L^* = 2$). Therefore, (2.103) becomes

$$c^{(f)}(\boldsymbol{n}, i) = E\left\{\frac{1}{\Delta}\left\{\int_0^{T'_{L^*}} f[\boldsymbol{N}'(t)]dt - \int_0^{T_{L^*}} f[\boldsymbol{N}(t)]dt\right\}\right\}. \quad (2.104)$$

Thus, $c^{(f)}(\boldsymbol{n}, i)$ defined in (2.103) is finite with probability 1 (cf. (2.5) for Markov chains).

Next, we study the effect of two or more perturbations at different servers. Consider a sample path of a closed network consisting of M servers. Suppose that at time $t = 0$, both server 1 and server 2 obtain a perturbation denoted as Δ_1 and Δ_2, respectively, with the same size $\Delta_1 = \Delta_2 = \Delta$. Let us propagate Δ_1 and Δ_2 separately along the sample path. First, we consider the propagation of Δ_1 at server 1. During the propagation, we use a 0-1 row vector $w_1(t)$ to denote which server has the perturbation at time $t \in [0, \infty)$. Specifically, we define $w_{1,i}(t) = 1$ if server i has the perturbation at time t, $w_{1,i}(t) = 0$ if otherwise, where $w_{1,i}(t)$ is the ith component of $w_1(t)$. Thus, initially the situation is represented by the vector $w_1(0) = (1, 0, 0, \ldots, 0)$. According to the propagation rules, when server i terminates an idle period of server j, server i's perturbation (either 0 or Δ) will be propagated to server j. This is equivalent to simply setting $w_{1,j} := w_{1,i}$ after the propagation.

Similarly, the propagation of the perturbation Δ_2 starts with the vector $w_2(0) = (0, 1, 0, \ldots, 0)$. We combine both vectors $w_1(0)$ and $w_2(0)$ together as an array

$$\begin{bmatrix} 1\,0\,0\,0 \ldots 0 \\ 0\,1\,0\,0 \ldots 0 \end{bmatrix}. \quad (2.105)$$

Now, let us propagate $\varDelta_1 (= \varDelta)$ and $\varDelta_2 (= \varDelta)$ simultaneously along the same sample path. As explained above, the propagation process is equivalent to copying the ith column of the above array to its jth column when server i terminates an idle period of server j. Thus, it is clear that, during propagation, the columns in the array (2.105) can never be $(1, 1)^T$. That is, if we propagate both perturbations $\varDelta_1$ and $\varDelta_2$ together along the same sample path, any transition time of this sample path can acquire at most one of the perturbations, never both. In other words, if, at any time, a server has a perturbation, then this perturbation is propagated from either $\varDelta_1$ or $\varDelta_2$. Eventually, the array may reach one of the following three situations:

$$\begin{bmatrix} 0\,0\dots 0 \\ 0\,0\dots 0 \end{bmatrix}, \quad \begin{bmatrix} 0\,0\dots 0 \\ 1\,1\dots 1 \end{bmatrix}, \quad \begin{bmatrix} 1\,1\dots 1 \\ 0\,0\dots 0 \end{bmatrix}.$$

That is, either one of them is realized, or both are lost, on the sample path; but they cannot be both realized. Furthermore, the propagation of one perturbation (say $\varDelta_1$) does not interfere (change) the propagation of the other (say $\varDelta_2$). That is, each perturbation is propagated along the sample path in the same way as if the other did not exist.

Based on this observation, we have the *superposition of the propagation of perturbations* on a sample path: If we propagate two perturbations of servers i and j, with the same size, simultaneously on a sample path $\boldsymbol{N}(t)$ and obtain a perturbation path $\boldsymbol{N}'(t)$, then we have

$$c^{(f)}(\boldsymbol{n}, i) + c^{(f)}(\boldsymbol{n}, j) = E\left\{ \frac{1}{\varDelta} \Big\{ \int_0^{T'_{L^*}} f[\boldsymbol{N}'(t)]dt - \int_0^{T_{L^*}} f[\boldsymbol{N}(t)]dt \Big\} \right\}.$$

The same discussion applies to the propagation of more than two perturbations. Let $V \subseteq \{1, 2, \dots, M\}$. Suppose that at time $t = 0$, all the servers in set V obtain the same perturbation $\varDelta_i = \varDelta$, $i \in V$. We propagate all these perturbations simultaneously on a sample path $\boldsymbol{N}(t)$ and obtain a perturbation path $\boldsymbol{N}'(t)$. Then, we have

$$\sum_{i \in V} c^{(f)}(\boldsymbol{n}, i) = E\left\{ \frac{1}{\varDelta} \left\{ \int_0^{T'_{L^*}} f[\boldsymbol{N}'(t)]dt - \int_0^{T_{L^*}} f[\boldsymbol{N}(t)]dt \right\} \right\}. \qquad (2.106)$$

Now we are ready to show that $c^{(f)}(\boldsymbol{n}, i)$ satisfy the following set of linear equations [43, 51].

1. If $n_i = 0$, then $c^{(f)}(\mathbf{n}, i) = 0$.
2. $\sum_{i=1}^{M} c^{(f)}(\mathbf{n}, i) = f(\mathbf{n})$.
3. Let $\mathbf{n}_{-i,+j} = (n_1, \dots, n_i - 1, \dots, n_j + 1, \dots, n_M)$ be a neighboring state of $\boldsymbol{n}$. Then,

$$\left[\sum_{i=1}^{M}\epsilon(n_i)\mu_i\right]c^{(f)}(\mathbf{n},k)=\sum_{i=1}^{M}\sum_{j=1}^{M}\epsilon(n_i)\mu_i q_{i,j}c^{(f)}(\mathbf{n}_{-i,+j},k)$$
$$+\sum_{j=1}^{M}\mu_k q_{k,j}\left\{[1-\epsilon(n_j)]c^{(f)}(\mathbf{n}_{-k,+j},j)+f(\mathbf{n})-f(\mathbf{n}_{-k,+j})\right\},$$
$$n_k>0,\quad k=1,2,\ldots,M, \tag{2.107}$$

where $\epsilon(n_j)=0$, if $n_j=0$, and $\epsilon(n_j)=1$, if $n_j>0$.

The above equations can be easily derived. Property 1 is simply a convention: When a server is idle, any perturbation will be lost with probability 1 because after the idle period the server's perturbation is determined by another server that does not have the perturbation. Property 2 is a direct consequence of the superposition of propagation (2.106): Set $V=\{1,2,\ldots,M\}$. By definition, this means that every server has the same perturbation Δ at $T_0=0$, hence $L^*=0$; i.e., $T'_L=T_L+\Delta$ for all $L\geq L^*=0$. In particular, $T_{L^*}=0$ and $T'_{L^*}=\Delta$. Therefore,

$$\begin{aligned}F'_L-F_L&=\int_0^{T'_L}f[\boldsymbol{N}'(t)]dt-\int_0^{T_L}f[\boldsymbol{N}(t)]dt\\&=\left\{\int_0^{T'_{L^*}}f[\boldsymbol{N}'(t)]dt-\int_0^{T_{L^*}}f[\boldsymbol{N}(t)]dt\right\}\\&\quad+\left\{\int_{T'_{L^*}}^{T'_L}f[\boldsymbol{N}'(t)]dt-\int_{T_{L^*}}^{T_L}f[\boldsymbol{N}(t)]dt\right\}\\&=\int_0^{\Delta}f[\boldsymbol{N}'(t)]dt=f(\boldsymbol{n})\Delta.\end{aligned}$$

This leads to the second property. Equation (2.107) can be derived by the theorem of total probability. In (2.107), we assume that server k has a perturbation. $\frac{\epsilon(n_i)\mu_i q_{i,j}}{\sum_{i=1}^{M}\epsilon(n_i)\mu_i}$ is the probability that the next transition is from server i to server j, $i,j=1,2,\ldots,M$. If no idle period is involved in this transition, there is no perturbation propagation and server k keeps the same perturbation after the transition except that the system state changes to $\boldsymbol{n}_{-i,+j}$. This is reflected by the first term on the right-hand side. If there is an idle period at server j (i.e., $1-\epsilon(n_j)=1$), then, in addition to the perturbation in server k, the perturbation will be propagated from server k to server j. This is reflected by the second term on the right-hand side. $f(\mathbf{n})-f(\mathbf{n}_{-k,+j})$ is the effect due to the delay of the transition from server k to server j. Equation (2.107) implies that the effect of a perturbation before a transition equals the weighted sum, by transition probabilities, of the effects of the perturbations after the transition, plus the effect due to the delay of the transition. It has been proved that (2.107) and the equations in Properties 1 and 2 have a unique solution for irreducible closed Jackson networks [51, 113].

From (2.104), if $f(\boldsymbol{n}) = I(\boldsymbol{n}) = 1$ for all $\boldsymbol{n} \in \mathcal{S}$, we have

$$c^{(I)}(\boldsymbol{n}, i) = E\left[\frac{T'_{L^*} - T_{L^*}}{\Delta}\right].$$

From the meaning of the realization probability, we have $E[T'_{L^*} - T_{L^*}] = c(\boldsymbol{n}, i)\Delta$. Thus, $c(\boldsymbol{n}, i) = c^{(I)}(\boldsymbol{n}, i)$. Therefore, the realization probabilities satisfy the following equations:

1. If $n_i = 0$, then $c(\mathbf{n}, i) = 0$.
2. $\sum_{i=1}^{M} c(\mathbf{n}, i) = 1$.
3. If $n_k > 0$, $k = 1, 2, \ldots, M$, then

$$\left[\sum_{i=1}^{M} \epsilon(n_i)\mu_i\right] c(\mathbf{n}, k) = \sum_{i=1}^{M}\sum_{j=1}^{M} \epsilon(n_i)\mu_i q_{i,j} c(\mathbf{n}_{-i,+j}, k) + \sum_{j=1}^{M} \mu_k q_{k,j}\{[1 - \epsilon(n_j)]c(\mathbf{n}_{-k,+j}, j)\}.$$

The following example taken from [51] provides some idea of the numerical values for the realization probabilities.

Example 2.8. Consider a closed Jackson network with $M = 3$, $N = 5$, $\bar{s}_1 = 10$, $\bar{s}_2 = 8$, $\bar{s}_3 = 5$, and routing probability matrix

$$Q = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.8 & 0 & 0.2 \\ 0.3 & 0.7 & 0 \end{bmatrix}.$$

The realization probabilities are obtained by solving the set of equations. The results, together with the steady-state probabilities, are listed in Table 2.9. □

2.4.3 Performance Derivatives

We have now quantified the effect of a single perturbation on the long-run average reward. Next, we will determine the effect of a small change in a mean service time. Suppose that the mean service time of server v changes from $\bar{s}_v$ to $\bar{s}_v + \Delta\bar{s}_v$. Let $s_{v,l}$, $l = 1, 2, \ldots$, be the service time of the lth customer served at server v. Following the perturbation generation rule (2.97), the lth customer's service completion time at server v will gain a perturbation $\Delta_{v,l} = s_{v,l}\frac{\Delta\bar{s}_v}{\bar{s}_v} = \kappa s_{v,l}$, $l = 1, 2, \ldots$. All these perturbations will be propagated along the sample path. To calculate the effect of a small change in the mean

n	π(**n**)	c(**n**, 1)	c(**n**, 2)	c(**n**, 3)
(5,0,0)	0.19047	1.00000	0.00000	0.00000
(4,0,1)	0.06644	0.90584	0.00000	0.09416
(4,1,0)	0.15061	0.89385	0.10615	0.00000
(3,0,2)	0.02318	0.78826	0.00000	0.21174
(3,1,1)	0.05254	0.77060	0.17336	0.05604
(3,2,0)	0.11908	0.74279	0.25721	0.00000
(2,0,3)	0.00809	0.62556	0.00000	0.37444
(2,1,2)	0.01833	0.60901	0.25029	0.14070
(2,2,1)	0.04154	0.58286	0.37574	0.04141
(2,3,0)	0.09416	0.54528	0.45472	0.00000
(1,0,4)	0.00282	0.38089	0.00000	0.61911
(1,1,3)	0.00639	0.37327	0.34810	0.27863
(1,2,2)	0.01449	0.35728	0.51926	0.12346
(1,3,1)	0.03285	0.33315	0.62079	0.04606
(1,4,0)	0.07445	0.29754	0.70246	0.00000
(0,0,5)	0.00098	0.00000	0.00000	1.00000
(0,1,4)	0.00223	0.00000	0.48951	0.51049
(0,2,3)	0.00505	0.00000	0.71510	0.28490
(0,3,2)	0.01146	0.00000	0.83485	0.16515
(0,4,1)	0.02597	0.00000	0.91819	0.08181
(0,5,0)	0.05887	0.00000	1.00000	0.00000

Table 2.9. A Numerical Example of Realization Probabilities

service time $\bar{s}_v$, we need to add up the effect of all these single perturbations on the system performance.

Let $\pi(\boldsymbol{n})$ be the steady-state probability of state $\boldsymbol{n}$. Consider a time period $[0, T_L]$ with $L >> 1$. The length of the total time when the system is in state $\boldsymbol{n}$ in $[0, T_L]$ is $T_L\pi(\boldsymbol{n})$. The total perturbation generated in this period at server v due to the change $\Delta\bar{s}_v$ in the mean service time is $T_L\pi(\boldsymbol{n})\frac{\Delta\bar{s}_v}{\bar{s}_v}$. Since each perturbation on average has an effect of $c^{(f)}(\boldsymbol{n}, v)$ on F_L, the overall effect on F_L of all the perturbations induced when the system state is $\boldsymbol{n}$ is $[T_L\pi(\boldsymbol{n})\frac{\Delta\bar{s}_v}{\bar{s}_v}]c^{(f)}(\boldsymbol{n}, v)$. Finally, the total effect of the mean service time change, $\Delta\bar{s}_v$, on F_L is

$$\Delta F_L \approx \sum_{\text{all } \boldsymbol{n}} T_L\pi(\boldsymbol{n})\frac{\Delta\bar{s}_v}{\bar{s}_v}c^{(f)}(\boldsymbol{n}, v).$$

From this, we have

$$\frac{\bar{s}_v}{T_L/L}\frac{\Delta F_L/L}{\Delta \bar{s}_v} \approx \sum_{\text{all } \boldsymbol{n}} \pi(\boldsymbol{n}) c^{(f)}(\boldsymbol{n}, v).$$

Letting $L \to \infty$ and then $\Delta \bar{s}_v \to 0$, we obtain the steady-state performance derivative as follows:

$$\boxed{\frac{\bar{s}_v}{\eta^{(I)}}\frac{\partial \eta^{(f)}}{\partial \bar{s}_v} = \sum_{\text{all } \boldsymbol{n}} \pi(\boldsymbol{n}) c^{(f)}(\boldsymbol{n}, v), \qquad (2.108)}$$

where $\eta^{(I)} = \lim_{L\to\infty} \frac{T_L}{L} = \frac{1}{\eta}$, see (2.95). Thus, the *normalized* derivative of the average reward (the left-hand side of (2.108)) equals the steady-state expectation of the realization factor. The above discussion provides an intuitive derivation and explanation for (2.108). See (2.116) in the next section for a formal formulation.

Set $f = I$ in (2.108). With $\eta^{(I)} = \frac{1}{\eta}$ and $c(\boldsymbol{n}, v) = c^{(I)}(\boldsymbol{n}, v)$, we can express the "elasticity" (normalized derivative) of the system throughput by using the perturbation realization probabilities:

$$\boxed{\frac{\bar{s}_v}{\eta}\frac{\partial \eta}{\partial \bar{s}_v} = -\sum_{\text{all } \boldsymbol{n}} \pi(\boldsymbol{n}) c(\boldsymbol{n}, v). \qquad (2.109)}$$

Summing up both sides over $v = 1, 2, \ldots, M$, we have

$$\sum_{v=1}^{M} \frac{\bar{s}_v}{\eta}\frac{\partial \eta}{\partial \bar{s}_v} = -1. \tag{2.110}$$

Example 2.9. In this example [51], we choose $M = 3$, $N = 8$, $\bar{s}_1 = 5$, $\bar{s}_2 = 10$, and $\bar{s}_3 = 12$. The routing probability matrix is

$$Q = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.7 & 0 & 0.3 \\ 0.4 & 0.6 & 0 \end{bmatrix}.$$

The realization probability equations are solved numerically. The elasticities calculated by (2.109) are -0.0365, -0.5133, and -0.4502, which are exactly the same as those calculated by queueing theory formulas. These values also satisfy (2.110). □

As shown in Section 2.4.1, the elasticity of the throughput can be estimated by a very efficient algorithm, Algorithm 2.1, together with equation (2.101). A close examination of the algorithm reveals that it, in fact, estimates the right-hand side of (2.109); i.e., it estimates the total sum as (cf. (2.101)):

$$\sum_{\text{all } \boldsymbol{n}} \pi(\boldsymbol{n}) c(\boldsymbol{n}, v) \approx \frac{\Delta_u}{T_L}.$$

Roughly speaking, $T_L \pi(\boldsymbol{n})$ is proportional to the perturbation generated when the system is in state $\boldsymbol{n}$; we may use $T_L \pi(\boldsymbol{n})$ as the perturbation generated, which corresponds to setting $\kappa = 1$ in Algorithm 2.1. At the end of the simulation, $\Delta_u \approx \sum_{\text{all } \boldsymbol{n}} T_L \pi(\boldsymbol{n}) c(\boldsymbol{n}, v)$ contains all the realized perturbations.

Similarly, with Algorithm 2.2, we, in fact, are estimating the performance derivative by

$$\sum_{\text{all } \boldsymbol{n}} \pi(\boldsymbol{n}) c^{(f)}(\boldsymbol{n}, v) \approx \frac{\Delta F}{T_L}.$$

Again, $T_L \pi(\boldsymbol{n})$ is proportional to the perturbation generated in state $\boldsymbol{n}$, and ΔF reflects the differences in performance realized due to all these perturbations.

Example 2.10. Consider the mean response time $\bar{\tau}$ in an M/G/1 queue in Example 2.4. Let $f = n$ be the number of customers in the server. We have $F_L = \int_0^{T_L} n(t) dt$, and $\bar{\tau} = \lim_{L \to \infty} \frac{F_L}{L}$. (In the definition of $\bar{\tau}$, L should be the number of departures. However, since the number of arrivals roughly equals that of the departures, we may take L be the number of all transitions, including both arrivals and departures, and the normalized derivative will be the same.) Suppose that the arrival rate does not change but the service rate changes. Then, the perturbation generation rule is (2.99), i.e., the perturbations of the service times are proportional to the service times. At a service completion time, the system state changes from n to $n-1$, $n > 0$, so $f(\boldsymbol{n}) - f(\boldsymbol{n}') = 1$ in Algorithm 2.2. The perturbations at the service completion times are calculated in (2.100). Thus, the perturbation of F_L calculated by Algorithm 2.2 is

$$\Delta F = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \sum_{l=1}^{i} s_{k,l},$$

where K is the number of busy periods in the L transitions, and n_k is the number of customers served in the kth busy period. (The real change in F_L should be $\kappa \Delta F_L$.) Finally, we have

$$\frac{\mu}{\eta^{(I)}} \frac{\partial \bar{\tau}}{\partial \mu} \approx -\frac{\Delta F}{T_L} = -\frac{1}{T_L} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \sum_{l=1}^{i} s_{k,l}. \qquad (2.111)$$

It can be proved that the right-hand side of (2.111) is indeed a strongly consistent estimate of its left-hand side (see the discussions in [103, 104, 146, 234, 235] and Problem 2.32). □

Comparison of PA of Queueing Systems and PA of Markov Chains

In PA of queueing systems, a small (*"infinitesimal"*, the exact meaning of this word will become clear in the next section) change in a system parameter (such as the mean service time of a server) induces a series of small (infinitesimal) changes of the state transition times on a sample path; each such change is called a perturbation (perturbation generation). These perturbations will be propagated along the sample path and affect the transition times of other state transitions (perturbation propagation). For irreducible networks, the effect of such a small perturbation on a sample path cannot continue forever; eventually, a perturbation will be either realized or lost on any sample path (perturbation realization). The average effect of each perturbation on the system performance can be precisely measured by a quantity called the perturbation realization factor (PRF). The total effect of a small change in a system parameter on the system performance can then be calculated by adding together the average effects of all the perturbations induced by the parameter change. The derivative of the performance with respect to the parameter can then be determined.

In PA of Markov chains, a small change in a system parameter (such as the transition probability matrix) induces a series of changes in the state transitions on a sample path; each such change is a perturbation and is also called a "jump" to intuitively reflect its discrete and finite nature. Thus, in PA of queueing systems a perturbation is an "infinitesimal" change on a sample path; while in PA of Markov chains, it is a finite change on a sample path. Moreover, perturbation propagation is not so distinct in PA of Markov chains, although we may view the Markov system as in a propagation period before the perturbed sample path merges with the original one. The perturbation realization principle and the calculation of performance derivatives for Markov chains are essentially the same as those for queueing systems: a single perturbation (jump) can only affect the system in a finite period (until the perturbed path merges with the original one), and its effect on the system performance can be measured by PRF, and so on.

In general, given a sample path of any system, we may first examine how a parameter change induces perturbations on a sample path (perturbation generation) and then determine how each perturbation affects the system performance (perturbation realization). During this process, we may explore how the system dynamics may help in determining the evolution of perturbations and whether there are simple propagation rules. These PA principles are illustrated in Figure 2.20. Again, this approach is of an intuitive nature and the results obtained need to be rigorously proved (cf. Section 2.4.4).

2.4.4 Remarks on Theoretical Issues*

The previous subsections provided an intuitive explanation for PA. The results have to be theoretically studied in a probability and statistical framework.

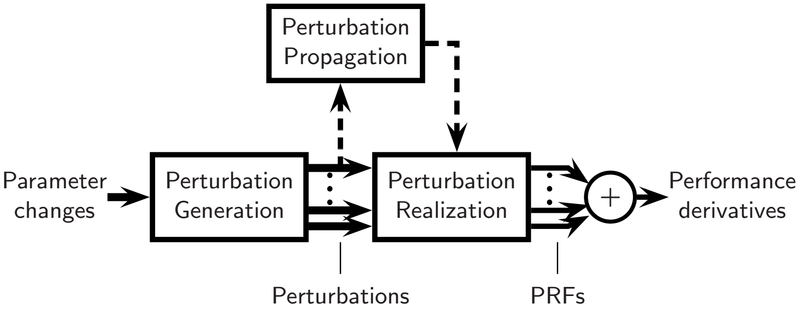


Fig. 2.20. The PA Principles

For example, if we use $-\frac{\Delta_u}{T_L}$ in (2.101) as an estimate for the elasticity of the throughput with respect to a mean service time, then is this estimate strongly consistent as L goes to infinity? Furthermore, is an estimate for performance derivative obtained in a finite sample path an unbiased estimate? These issues were first formulated and studied in [42], and later they were studied for many problems by many authors, and [51] provides a detailed summary of the theory. It is out of the scope of this book to discuss all these issues in detail and we give only a brief review here.

Sample Functions and Sample Derivatives

First let us mathematically describe a sample path of a closed Jackson network obtained by simulation. With the inverse-transform method (2.96), the lth service time at server i, $s_{i,l}$, can be obtained by a uniformly distributed random number $\xi_{i,l} \in [0,1)$ with $s_{i,l} = -\bar{s}_i \ln(1-\xi_{i,l})$. Therefore, all the service times on a sample path depend on a sequence of random numbers $\{\xi_{1,1}, \xi_{1,2}, \ldots; \xi_{2,1}, \xi_{2,2}, \ldots; \ldots; \xi_{M,1}, \xi_{M,2}, \ldots\}$; they are independent and uniformly distributed on $[0,1)$. After the completion of its service, a customer at server i will move to server j, $j = 1, 2, \ldots, M$, with probability $q_{i,j}$. Thus, the next destination of the lth customer at server i can be determined by another uniformly distributed random number $\zeta_{i,l} \in [0,1)$: if $\sum_{k=1}^{j-1} q_{i,k} \le \zeta_{i,l} < \sum_{k=1}^{j} q_{i,k}$ (with the convention $\sum_{k=1}^{0} q_{i,k} = 0$), then this customer moves to server j. Therefore, all the destinations depend on another sequence of independent and uniformly distributed $[0,1)$ random numbers $\{\zeta_{1,1}, \zeta_{1,2}, \ldots; \zeta_{2,1}, \xi_{2,2}, \ldots; \ldots; \zeta_{M,1}, \zeta_{M,2}, \ldots\}$. Finally, let $\xi = \{\xi_{1,1}, \xi_{1,2}, \ldots; \ldots; \xi_{M,1}, \xi_{M,2}, \ldots; \zeta_{1,1}, \zeta_{1,2}, \ldots; \ldots; \zeta_{M,1}, \zeta_{M,2}, \ldots\}$. Then, ξ represents all the randomness involved in the system. Let $\theta = \{\bar{s}_i, q_{i,j}, i, j = 1, 2, \ldots, M\}$ represent all the parameters in the system. With these notations, a sample path of the system is determined by, and therefore is denoted as, (ξ, θ).

For any fixed integer L, F_L in (2.94) and $\eta_L^{(f)} = \frac{F_L}{L}$ are defined on a sample path and therefore are functions of (ξ, θ). We denote them as $F_L(\xi, \theta)$

and $\eta_L^{(f)}(\xi, \theta)$. As we can see from (2.96)-(2.97), in a perturbed sample path, the same sequence of random numbers ξ is used, but the parameters may experience a small change. Thus, a perturbed sample path is in fact $(\xi, \theta + \Delta\theta)$. The perturbed performance is $F_L(\xi, \theta + \Delta\theta)$. Of course, $\Delta\theta$ may be zero for many of its components. In the Jackson network studied in this section, we only choose $\Delta\bar{s}_v \neq 0$ for server v. Since we are concerned with the performance derivatives, for notational simplicity, let us assume that θ is a scalar parameter that changes to $\theta + \Delta\theta$, $\Delta\theta \neq 0$.

The perturbation generation and propagation rules help us to construct the perturbed sample path $(\xi, \theta + \Delta\theta)$ from the original sample path (ξ, θ) for a small $\Delta\theta$ (by Algorithm 2.1), and then to obtain the perturbed performance $F_L(\xi, \theta + \Delta\theta)$ (by Algorithm 2.2). We have

$$\Delta F_L(\xi, \theta) = F_L(\xi, \theta + \Delta\theta) - F_L(\xi, \theta)$$

and

$$\Delta\eta_L^{(f)}(\xi, \theta) = \frac{1}{L}[F_L(\xi, \theta + \Delta\theta) - F_L(\xi, \theta)]. \tag{2.112}$$

For any fixed ξ, $\eta_L^{(f)}(\xi, \theta)$ or $F_L(\xi, \theta)$ is a function of θ. We call it a *sample performance function* [46, 51].

When we apply the propagation rules, we require the perturbation of any server, Δ, to be small enough. In fact, Δ should be smaller than the length of an idle period in order for the perturbation Δ to be propagated through the idle period without changing its size. Figure 2.21 shows the situation when a perturbation is larger than an idle period. Figure 2.21.A illustrates the same sample path as Figure 2.15, except that the perturbation Δ_1 is larger than the length of the idle period $T_2 - T_1$. Figure 2.21.B illustrates the corresponding perturbed path. Indeed, when Δ_1 is larger than $T_2 - T_1$, the idle period in server 1 disappears in the perturbed path and a new idle period appears in server 2. The order of the transition times of server 1 and server 2 changes: $T_2 > T_1$ in the original path, but $T_1' > T_2'$ in the perturbed one. Both servers are delayed by $\Delta_1 - (T_2 - T_1)$ after the idle period. All these facts indicate that the simple propagation rules used in Algorithm 2.1 do not apply.

In fact, Algorithm 2.2 requires a more strict condition: the perturbation of any server in $[0, T_L)$ should be smaller than the shortest sojourn time of the system in any state in $[0, T_L)$. For any finite L and a fixed sample path (ξ, θ), we can always choose (with probability 1) $\Delta\theta$ to be small enough such that this requirement is satisfied (this explains the meaning of infinitesimal). Thus, PA Algorithm 2.2 provides the exact value of $\Delta\eta_L^{(f)}(\xi, \theta)$ in (2.112) if $\Delta\theta$ is small enough. That is, what we obtained from PA is in fact the derivative of a sample performance function, which is called a *sample derivative* [46, 51]:

$$\boxed{\frac{\partial\eta_L^{(f)}(\xi, \theta)}{\partial\theta} = \lim_{\Delta\theta \to 0} \frac{\eta_L^{(f)}(\xi, \theta + \Delta\theta) - \eta_L^{(f)}(\xi, \theta)}{\Delta\theta}, \qquad \text{for a fixed } \xi.}$$

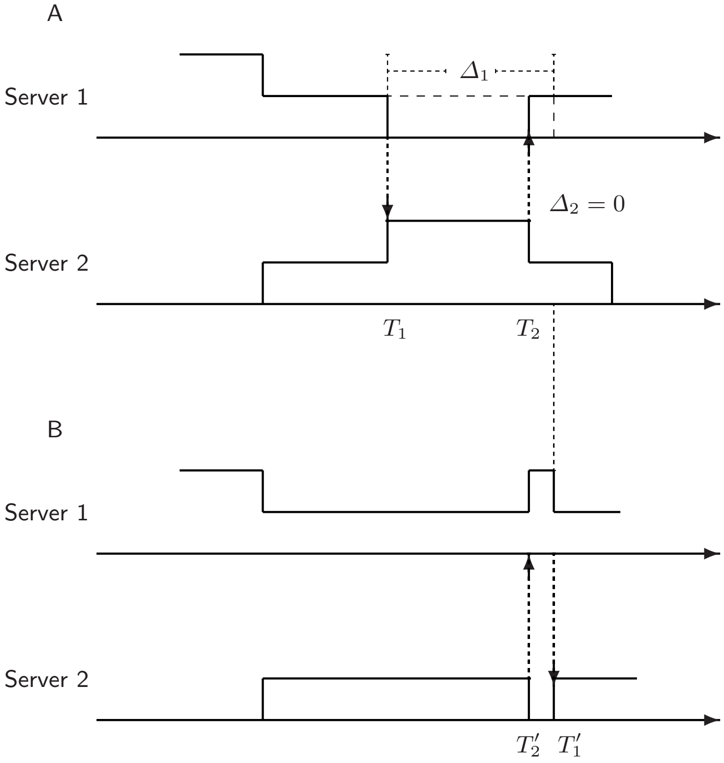


Fig. 2.21. A Large Perturbation Does Not Satisfy the Propagation Rule

Interchangeability

In general, however, we are interested in the derivative of the mean performance $E[\eta_L^{(f)}(\xi,\theta)]$, $\frac{\partial E[\eta_L^{(f)}(\xi,\theta)]}{\partial\theta}$, or the derivative of the steady-state performance $\eta^{(f)}(\theta) = \lim_{L\to\infty} \eta_L^{(f)}(\xi,\theta)$, $\frac{\partial\eta^{(f)}(\theta)}{\partial\theta}$. This raises two questions: Is the sample derivative obtained by PA on a sample path in a finite period $[0, T_L)$ an unbiased estimate? That is, for any $L < \infty$, does

$$E\left\{\frac{\partial}{\partial\theta}[\eta_L^{(f)}(\xi,\theta)]\right\} = \frac{\partial}{\partial\theta}\left\{E[\eta_L^{(f)}(\xi,\theta)]\right\}? \qquad (2.113)$$

Also, is it a strong consistent estimate? That is, does

$$\lim_{L\to\infty}\left\{\frac{\partial}{\partial\theta}[\eta_L^{(f)}(\xi,\theta)]\right\} = \frac{\partial\eta^{(f)}}{\partial\theta} = \frac{\partial}{\partial\theta}\left\{\lim_{L\to\infty}[\eta_L^{(f)}(\xi,\theta)]\right\}? \qquad (2.114)$$

In calculus, (2.113) or (2.114) means that the order of the two operators "E" and "$\frac{\partial}{\partial\theta}$", or "$E$" and "$\lim_{L\to\infty}$", is interchangeable. This interchangeability requires some conditions on the sample performance function. The following simple example gives some ideas about why such interchangeability may not hold for some systems.

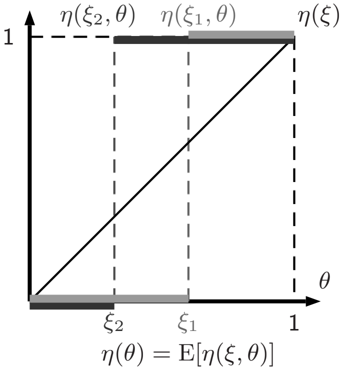


Fig. 2.22. A Sample Function That Does Not Satisfy Interchangeability

Example 2.11. Consider a sample function defined as

$$\eta(\xi,\theta) = \begin{cases} 1, & \text{if } \theta > \xi, \\ 0, & \text{otherwise,} \end{cases} \qquad (2.115)$$

where ξ is a uniformly distributed random variable in $[0,1)$. η equals 1 if $\theta \in [\xi, 1)$ and 0 if $\theta \in [0, \xi)$. Two such sample paths corresponding to ξ_1 and ξ_2 are illustrated in Figure 2.22. The mean performance is $\eta(\theta) = E[\eta(\xi,\theta)] = \theta$. The sample derivative is the slope of the sample function $\eta(\xi,\theta)$, which equals 0 with probability 1. Therefore, we have

$$E\left\{\frac{\partial}{\partial\theta}[\eta(\xi,\theta)]\right\} = 0 \neq \frac{\partial}{\partial\theta}\{E[\eta(\xi,\theta)]\} = 1.$$

That is, the interchangeability does not hold for this sample function. □

Fortunately, we can prove that for closed Jackson networks with any finite reward function $f(\boldsymbol{n})$, $\boldsymbol{n} \in \mathcal{S}$, it does hold [46, 51]

$$E\left\{ \frac{\partial}{\partial \bar{s}_v} \left[\eta_L^{(f)}(\xi, \bar{s}_v)\right] \middle| X_0 = \boldsymbol{n}_0 \right\} = \frac{\partial}{\partial \bar{s}_v} E\left\{ \left[\eta_L^{(f)}(\xi, \bar{s}_v)\right] \middle| X_0 = \boldsymbol{n}_0 \right\},$$

where $\boldsymbol{n}_0$ is any initial state. This equation shows that the sample derivative provided by PA, $\frac{\partial}{\partial \bar{s}_v}[\eta_L^{(f)}(\xi, \bar{s}_v)]$, is unbiased for the derivative of the mean (transient) average reward in $[0, T_L)$. In particular, when $f \equiv I$, we have [49]

$$E\left\{ \frac{\partial}{\partial \bar{s}_v}[T_L(\xi, \bar{s}_v)] \middle| X_0 = \boldsymbol{n}_0 \right\} = \frac{\partial}{\partial \bar{s}_v} E\{[T_L(\xi, \bar{s}_v)] | X_0 = \boldsymbol{n}_0\}.$$

For long-run average rewards, we also have [51]

$$\begin{aligned} \lim_{L \to \infty} \left[\frac{\bar{s}_v}{\eta_L^{(I)}(\xi, \bar{s}_v)} \frac{\partial \eta_L^{(f)}(\xi, \bar{s}_v)}{\partial \bar{s}_v} \right] &= \frac{\bar{s}_v}{\eta^{(I)}(\bar{s}_v)} \frac{\partial \eta^{(f)}(\bar{s}_v)}{\partial \bar{s}_v} \\ &= \sum_{\text{all } \boldsymbol{n}} \pi(\boldsymbol{n}) c^{(f)}(\boldsymbol{n}, v), \qquad \text{w.p.1}, \end{aligned} \tag{2.116}$$

where $\eta^{(f)}(\bar{s}_v) = \lim_{L \to \infty} \eta_L^{(f)}(\xi, \bar{s}_v)$ and $\eta^{(I)}(\bar{s}_v) = \lim_{L \to \infty} \eta_L^{(I)}(\xi, \bar{s}_v)$. In particular, we have

$$\begin{aligned} \lim_{L \to \infty} \left[\frac{\bar{s}_v}{\eta_L(\xi, \bar{s}_v)} \frac{\partial \eta_L(\xi, \bar{s}_v)}{\partial \bar{s}_v} \right] &= \frac{\bar{s}_v}{\eta(\bar{s}_v)} \frac{\partial \eta(\bar{s}_v)}{\partial \bar{s}_v} \\ &= -\sum_{\text{all } \boldsymbol{n}} \pi(\boldsymbol{n}) c(\boldsymbol{n}, v), \qquad \text{w.p.1}, \end{aligned} \tag{2.117}$$

where $\eta(\bar{s}_v) = \lim_{L \to \infty} \eta_L(\xi, \bar{s}_v)$, and $\eta_L(\xi, \bar{s}_v) = \frac{L}{T_L(\xi, \bar{s}_v)}$. That is, the normalized sample derivatives provided by PA are strongly consistent estimates of the normalized derivatives of the steady-state performance.

However, the nice properties of unbiasedness and strong consistency do not always hold. As illustrated in Example 2.11, the interchangeability may not hold if the sample functions are discontinuous. Roughly speaking, the interchangeability in (2.113) requires that the sample performance functions be "smooth" enough.

The sample derivatives of the performance with respect to the changes in routing probabilities $q_{i,j}$, $i, j = 1, 2, \ldots, M$, have discontinuities similar to those in Example 2.11. To demonstrate the idea, we consider a closed network and assume that its service time distributions do not change. A sample path of such a network is determined by the random variables $\zeta := \{\zeta_{1,1}, \zeta_{1,2}, \ldots; \zeta_{2,1}, \xi_{2,2}, \ldots; \ldots; \zeta_{M,1}, \zeta_{M,2}, \ldots\}$, and therefore we may denote a sample path as $(\zeta, q_{i,j}, i, j = 1, 2, \ldots, M)$. For the sake of discussion, we assume that $q_{1,2}$ and $q_{1,3}$ change to $q'_{1,2} = q_{1,2} + \delta$ and $q'_{1,3} = q_{1,3} - \delta$, respectively. As we know, in simulation, the customer transition is determined as follows: we first divide the interval $[0, 1]$ into M small segments, each with length $q_{1,i}$, $i = 1, 2, \ldots, M$ (see Figure 2.23 for $M = 3$). If $\zeta_{1,l}$ falls in the kth segment, then the lth customer at server 1 moves to server k. When $q_{1,2}$

and $q_{1,3}$ change to $q'_{1,2}$ and $q'_{1,3}$, respectively, the only change happens when $\zeta_{1,l}$ falls in the small segment with length δ (in the middle of the period $[0,1]$ shown in Figure 2.23). In this case, the customer moves to server 3 in the original sample path but to server 2 in the perturbed path. It is clear that for a fixed realization of ζ and a finite L, there is always (with probability 1) a δ_0 that is small enough such that no $\zeta_{1,l}$, $l = 1, 2, \ldots, L$, falls in that small segment. Therefore, the two sample paths $(\zeta, q_{i,j})$ and $(\zeta, q'_{i,j})$ with $\delta < \delta_0$ are exactly the same in $[0, T_L]$. This implies that the sample function $F_L(\zeta, q_{i,j})$ is a piecewise constant function of $q_{i,j}$. As shown in Example 2.11, the interchangeability in (2.113) does not hold for the derivative of performance $F_L(\zeta, q_{i,j})$ with respect to δ (or the changes in $q_{1,2}$ and $q_{1,3}$).

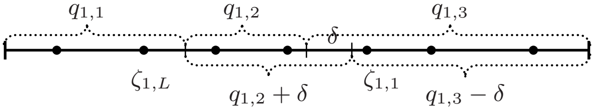


Fig. 2.23. Determine the Customer Transitions

The PA of queueing systems introduced in this section is based on sample derivatives. This approach requires interchangeability, which may not hold if the sample function is not continuous. The discontinuity of a sample function can be explained from a sample path point of view. Essentially, if a small change in a parameter may cause a big change in a sample path, the sample function may be discontinuous. In the case with the routing probabilities, a small change in $q_{1,2}$ (or $q_{1,3}$) may cause a big change in a customer's destination (from server 2 to server 3). Such a big change also occurs when two transitions exchange their order of occurrence, leading to two different states. This sample-path-based explanation gives us an intuitive feeling about whether the discontinuity may exist (see [42] and [126] for more details).

Other examples where the interchangeability does not hold include queueing networks with multi-class customers or with blocking due to finite buffer sizes. They can also be explained by the intuitive explanation described above (see [43] and [126] for more discussion).

For the same reason, the sample performance functions for Markov systems with respect to the transition probability matrix are also piecewise linear, and the sample-derivative is therefore zero and the approach discussed in this section does not apply. However, as shown in Section 2.1, the basic principle of perturbation generation and perturbation realization can be extended to Markov systems. The derivative obtained by using realization factors for Markov systems is not a sample derivative.

Similar results regarding the sample functions and sample derivatives for systems with continuous state spaces are presented in [47], and a comparison of the dynamics of the continuous and discrete event systems is given in [48].

2.5 Other Methods*

Much effort was expended in the 1980's to overcome the difficulty caused by the discontinuity of the sample functions for some systems. Different approaches were developed; these approaches work well for some special problems. Among them are the *smoothed perturbation analysis (SPA)* [105, 107, 114, 119], the *finite perturbation analysis (FPA)* [143], the *rare perturbation analysis (RPA)* [33, 34, 36, 37]. There are also other related works [73, 94, 102, 106, 108, 128, 147, 156, 185, 214, 232, 241, 245, 247, 251]. These topics have been widely discussed in previous books [51, 72, 107, 112, 142], and therefore we will not discuss them in this book.

In this section, we will briefly review some other methods of performance sensitivity analysis. They are the *stochastic fluid model*, the *weak derivative method*, and the *likelihood ratio* or *score function method*.

The Stochastic Fluid Model (SFM)

The stochastic fluid model (SFM) has been recently adopted to model complex, discrete-event dynamic systems such as communication networks, and perturbation analysis has been proposed in SFM as a means for sensitivity analysis. The essential idea of this method is to use a continuous flow to approximately model the packet transmission in a network. Since in communication a data or voice packet consists of small units called bits, SFM is particularly suitable for communication systems.

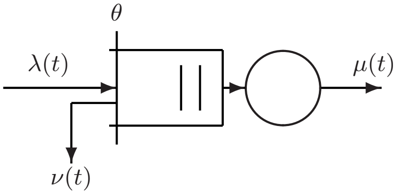


Fig. 2.24. The Stochastic Fluid Model for a Single Queue with Buffer Size θ

Figure 2.24 illustrates a stochastic fluid model for a single queue. The inflow rate and the processing rate at time t are denoted as $\lambda(t)$ and $\mu(t)$ (units/second), respectively; and we use θ (units) to denote the size of the buffer. When the buffer is full, the incoming fluid will overflow, and we denote its rate at $\nu(t)$. Let $x_\theta(t)$ be the volume of the fluid in the buffer. Apparently, the system dynamic can be modelled by

$$\frac{dx_\theta(t)}{dt} = \begin{cases} 0, & \text{if } x_\theta(t) = 0, \text{ and } \lambda(t) \le \mu(t), \\ 0, & \text{if } x_\theta(t) = \theta \text{ and } \lambda(t) > \mu(t), \\ \lambda(t) - \mu(t), & \text{otherwise.} \end{cases}$$

Sample-path-based analysis can be applied to such a SFM to obtain an estimate for the performance derivative. The approach is more suitable (although approximate) for estimating the gradients of packet loss probability with respect to the buffer size. It can be shown that such estimates are unbiased for the derivatives of the performance obtained with the SFM model. Such problems are usually difficult to handle with the standard PA of queueing systems. For more details and applications, see [74, 75, 189, 210, 211, 231, 252, 262, 263].

The Likelihood Ratio (Score Function) Method

Another performance derivative estimation method is called the *likelihood ratio* method, [44, 115, 116, 117, 118, 130, 176, 177, 178, 179, 205, 217], also called the *score function* method [221, 222].

To illustrate the main idea, let us consider a D/M/1 queue in which the inter-arrival time is a fixed number $D > 0$ and the service times are independent and exponentially distributed with mean $\bar{s}$. Let $s_1, s_2, \ldots, s_L$ be the sequence of customers' service times. Then, a sample path of the system can be represented by, and therefore denoted as, a vector $s := (s_1, s_2, \ldots, s_L)$ (instead of in the form of (ξ, θ)). The performance defined on this sample path is denoted as $\eta(s)$. Let $\varPhi(s, \theta)$ be the distribution function of s, where $\theta = \bar{s}$ denotes the system parameter. (To help our understanding, we may view s as a scalar variable, otherwise, $\varPhi(s, \theta)$ is the joint distribution of $s_1, \ldots, s_L$.) The mean performance is

$$\eta_\theta = E[\eta(s)] = \int_{-\infty}^{\infty} \eta(s) d\varPhi(s, \theta). \tag{2.118}$$

Our goal is to estimate the derivative $\frac{d\eta_\theta}{d\theta}$.

In PA, we set $\xi := \varPhi(s, \theta)$ to be a $[0, 1)$ uniformly distributed random variable. Then, we have $s = \varPhi^{-1}(\xi, \theta)$, and, for notational convenience, we denote it as $s = \varPhi^{-1}(\xi, \theta) := s(\xi, \theta)$. Thus, we have

$$\eta_\theta = \int_0^1 \eta[s(\xi, \theta)] d\xi.$$

As explained in Section 2.4.4, for any fixed $\xi \in [0, 1)$, $\eta[s(\xi, \theta)]$ is called a sample performance function. In PA, we use the sample derivative $\frac{d}{d\theta}\eta[s(\xi, \theta)]$ as an estimate of $\frac{d\eta_\theta}{d\theta}$. The issue is whether or not this estimate is unbiased, i.e., whether or not (cf. (2.113))

$$\frac{d\eta_\theta}{dt} = \frac{d}{d\theta}\left\{\int_0^1 \eta[s(\xi, \theta)] d\xi\right\} = \int_0^1 \frac{d}{d\theta}\{\eta[s(\xi, \theta)]\} d\xi?$$

In the sample derivative $\frac{d}{d\theta}\eta[s(\xi, \theta)]$, the same random variable ξ is used for both $\eta[s(\xi, \theta)]$ and $\eta[s(\xi, \theta + \varDelta\theta)]$. Thus, PA is also called it a *common random number (CRN)* method. It is known that using the common random number leads to the smallest variance in estimating the difference between two random variables (See Problem A.4). Therefore, the sample derivative usually has a small variance.

The rationale of the likelihood ratios method is as follows: Suppose that the probability density function of $\Phi(s,\theta)$ exists and denote it as $\phi(s,\theta) = \frac{d}{ds}\Phi(s,\theta)$. Then, (2.118) becomes

$$\eta(\theta) = \int_{-\infty}^{\infty} \eta(s)\phi(s,\theta)ds,$$

and we have, assuming that the two operators $\int$ and $\frac{d}{d\theta}$ can change their order,

$$\begin{aligned} \frac{d\eta_\theta}{d\theta} &= \int_{-\infty}^{\infty} \eta(s)\frac{d\phi(s,\theta)}{d\theta}ds \\ &= \int_{-\infty}^{\infty} \eta(s)\frac{d\phi(s,\theta)}{d\theta}ds \\ &= \int_{-\infty}^{\infty} \eta(s)\frac{d\ln[\phi(s,\theta)]}{d\theta}d\Phi(s,\theta) \\ &= E\left\{\eta(s)\frac{d\ln[\phi(s,\theta)]}{d\theta}\right\}. \end{aligned} \tag{2.119}$$

This indicates that we may use

$$\eta(s)\frac{d\ln[\phi(s,\theta)]}{d\theta} \tag{2.120}$$

as an unbiased estimate of the performance derivative $\frac{d\eta_\theta}{d\theta}$. In (2.120), we have

$$\frac{d\ln[\phi(s,\theta)]}{d\theta} = \frac{1}{\phi(s,\theta)}\frac{d[\phi(s,\theta)]}{d\theta}.$$

Observe that

$$\eta(s)\frac{d\ln[\phi(s,\theta)]}{d\theta} = \lim_{\Delta\theta\to 0}\frac{1}{\Delta\theta}\left\{\eta[s(\xi,\theta)]\frac{\phi(s,\theta+\Delta\theta)}{\phi(s,\theta)} - \eta[s(\xi,\theta)]\right\}. \tag{2.121}$$

Therefore, in the LR estimate (2.120), we in fact use

$$\eta[s(\xi,\theta)]\frac{\phi(s,\theta+\Delta\theta)}{\phi(s,\theta)}$$

in the place of $\eta[s(\xi,\theta+\Delta\theta)]$. The reason is that if the system parameter changes from θ to $\theta+\Delta\theta$, the same sample path s, and therefore the same sample performance value $\eta(s)$, will still be observed, but with a different probability that is adjusted by the likelihood ratio

$$\frac{\phi(s, \theta + \Delta\theta)}{\phi(s, \theta)}$$

(see [44] for more discussion). Therefore, this approach is called the *likelihood ratio (LR)*, or the *score function (SF)* method.

From (2.121), the LR estimate essentially uses the same sample path s as a possible realization of the system behavior under parameters $\theta + \Delta\theta$ and adjusts the probability of this sample path; hence, the LR method is also called the *common realization (CR)* method.

LR only requires the interchangeability of $\int$ and $\frac{d}{d\theta}$ to hold for the probability density function, which is usually smoother than the sample performance function (the s in $\eta(s)$ in (2.119) is fixed). Thus, an LR estimate is unbiased more often than a PA estimate is. However, the variance may be too large to be applicable [44]. Variance reduction techniques based on regenerative periods have been developed.

The Weak Derivative Method

In the weak derivative method [130, 132, 134], the derivative of the probability density function is expressed by the difference between two properly chosen probability density functions, and the performance derivative is then expressed by the difference between two expected values. For example, in (2.119), if we have $c(\theta) > 0$ and two density functions $\phi_1(s, \theta)$ and $\phi_2(s, \theta)$ such that

$$\frac{d\phi(s, \theta)}{d\theta} = c(\theta)[\phi_1(s, \theta) - \phi_2(s, \theta)]. \tag{2.122}$$

Then,

$$\frac{d\eta(\theta)}{d\theta} = c(\theta) \left[\int_{-\infty}^{\infty} \eta(s)\phi_1(s, \theta) d\theta - \int_{-\infty}^{\infty} \eta(s)\phi_2(s, \theta) d\theta \right],$$

which is the difference between the mean performance of two sample paths, one with probability density function $\phi_1(s, \theta)$ and the other with $\phi_2(s, \theta)$. The triple $(c(\theta), \phi_1(s, \theta), \phi_2(s, \theta))$ is called a *weak derivative* of $\phi(s, \theta)$. Obviously, it is not unique.

The same principle applies to the performance derivatives of Markov chains. Consider two Markov chains defined on the same state space $\mathcal{S} = \{1, 2, \ldots, S\}$ with two ergodic transition probability matrices P and P' and the same reward function f. Let $\Delta P = P' - P$, $P_\delta = P + \delta\Delta P$. We start with (2.23). From (2.13) and (2.14), the directional derivative along ΔP is

$$\frac{d\eta_\delta}{d\delta} = \pi \Delta P \sum_{l=0}^{\infty} (P^l - e\pi) f. \tag{2.123}$$

Corresponding to (2.122), we have

$$\frac{dP_\delta}{d\delta} = \Delta P = C(P^+ - P^-), \tag{2.124}$$

where P^+, P^-, and C are defined as follows: C is a diagonal matrix with nonzero diagonal components $c(i)$, $i = 1, 2, \dots, S$,

$$c(i) = \sum_{j=1}^{S} \max\{\Delta p(j|i), 0\},$$

$\Delta p(j|i) = p'(j|i) - p(j|i)$, $i, j = 1, 2, \dots, S$, and

$$p^+(j|i) = \begin{cases} \frac{1}{c(i)} \max\{\Delta p(j|i), 0\}, & \text{if } c(i) > 0, \\ 0, & \text{if } c(i) = 0; \end{cases}$$

$$p^-(j|i) = \begin{cases} \frac{1}{c(i)} \max\{-\Delta p(j|i), 0\}, & \text{if } c(i) > 0, \\ 0, & \text{if } c(i) = 0. \end{cases}$$

When $c(i) \neq 0$, the ith rows of P^+ and P^- are transition probability vectors; and when $c(i) = 0$, the ith rows of P^+ and P^- are zero. The triple (C, P^+, P^-) is called a weak derivative of P_δ. The decomposition of (2.124) is not unique, and there may be other weak derivatives of P_δ.

From (2.124) and $(\Delta P)e = 0$, the derivative (2.123) becomes

$$\begin{aligned} \frac{d\eta_\delta}{d\delta} &= \pi \Delta P \sum_{l=0}^{\infty} P^l f \\ &= \pi C(P^+ - P^-) \sum_{l=0}^{\infty} P^l f \\ &= \sum_{i=1}^{S} \pi(i) c(i) \sum_{l=0}^{\infty} (p_i^+ P^l f - p_i^- P^l f), \end{aligned} \tag{2.125}$$

where p_i^+ and p_i^- denote the ith rows of P^+ and P^-, respectively.

There is a sample-path-based interpretation of (2.125). Let $\boldsymbol{X}^+ = \{X_l^+, l = 0, 1, \dots\}$ be a Markov chain obtained as follows: Suppose that $X_0^+ = i$ is the initial state, and the first transition from X_0^+ to X_1^+ follows transition probability vector p_i^+, and the rest of the transitions at $l = 1, 2, \dots$ follow transition probability matrix P. Let $\boldsymbol{X}^-$ be a similar Markov chain except that the first transition from X_0^- to X_1^- follows p_i^-, with $X_0^- = i$. From (2.125), we have

$$\frac{d\eta_\delta}{d\delta} = \sum_{i=1}^{S} \pi(i) c(i) \sum_{l=0}^{\infty} E\{[f(X_l^+) - f(X_l^-)] | X_0^+ = X_0^- = i\}. \tag{2.126}$$

Therefore, the performance derivative can be expressed via the difference between two expectations on two different Markov chains $\boldsymbol{X}^+$ and $\boldsymbol{X}^-$. Furthermore, by the strong Markov property, the infinite sum $\sum_{l=0}^{\infty}$ can be replaced by a finite one $\sum_{l=0}^{L_{+,-}}$; at $L_{+,-}$, the two sample paths $\boldsymbol{X}^+$ and $\boldsymbol{X}^-$ merge together.

The form of (2.126) resembles the performance realization factors. In fact, from (2.126) we can easily derive (see [130])

$$\frac{d\eta_\delta}{d\delta} = \sum_{i=1}^{S} \pi(i)c(i) \left[\sum_{j_1,j_2=1}^{S} \gamma(j_2, j_1)p^+(j_1|i)p^-(j_2|i) \right]. \tag{2.127}$$

PROBLEMS

2.1. In Figure 2.2, the three points P_0, P_1, and P_2 represent three policies. Every point P in the triangle with these three points as vertices represents a randomized policy denoted as $P(\delta_1, \delta_1, \delta_2) = \delta_0 P_0 + \delta_1 P_1 + \delta_2 P_2$, $\delta_0 + \delta_1 + \delta_2 = 1$, with $P_0 = P(1, 0, 0)$, $P_1 = P(0, 1, 0)$, and $P_2 = P(0, 0, 1)$.

a. Determine the values of δ_0, δ_1, and δ_2 by the lengths of the segments shown in the figure.

b. Along the line from P_0 to P_1, we have the randomized policies $P_\delta = (1 - \delta)P_0 + \delta P_1$, $0 < \delta < 1$, and we can obtain the directional derivative in this direction, denoted as $\frac{d\eta_\delta}{d\delta}|_{P_0-P_1}$. Similarly, we can obtain the directional derivative in the direction from P_0 to P_2, denoted as $\frac{d\eta_\delta}{d\delta}|_{P_0-P_2}$. What is the directional derivative from P_0 to P? Express it in terms of $\frac{d\eta_\delta}{d\delta}|_{P_0-P_1}$ and $\frac{d\eta_\delta}{d\delta}|_{P_0-P_2}$. (*Hint: Along this direction, δ_1/δ_2 is fixed.*)

2.2. (Random walk) A random walker moves among five positions $i = 1, 2, 3, 4, 5$. At position $i = 2, 3, 4$, s/he moves to positions $i - 1$ and $i + 1$ with an equal probability $p(i - 1|i) = p(i + 1|i) = 0.5$; at the boundary positions $i = 1$ and $i = 5$, s/he bounces back with probability 1 $p(4|5) = p(2|1) = 1$. We are given a sequence of 20 $[0, 1)$-uniformly and independently distributed random variables as follows:

0.740, 0.605, 0.234, 0.342, 0.629, 0.965, 0.364, 0.230, 0.599, 0.079,
0.782, 0.219, 0.475, 0.051, 0.596, 0.850, 0.865, 0.434, 0.617, 0.969.

a. With this sequence, construct a sample path $\boldsymbol{X}$ of the random walk from X_0 to X_{20} according to (2.2). Set $X_0 = 3$.

b. Suppose that the perturbed transition probabilities are $p'(i - 1|i) = 0.3$, $p'(i + 1|i) = 0.7$, $i = 2, 3, 4$, and $p'(4|5) = p'(2|1) = 1$. Set $p_\delta(j|i) = p(j|i) + \delta[p'(j|i) - p(j|i)]$. By using the original sample path obtained in (a), construct a perturbed sample path $\boldsymbol{X}_\delta$, $\delta = 1$, following Figure 2.5. Use the following $[0, 1)$-uniformly and independently distributed random variables when $\boldsymbol{X}_\delta$ is different from $\boldsymbol{X}$ (use the lth number to determine the lth transition of $\boldsymbol{X}_\delta$, if $X_{\delta,l} \neq X_l$):

0.173, 0.086, 0.393, 0.804, 0.011, 0.233, 0.934, 0.230, 0.786, 0.410,
0.119, 0.634, 0.862, 0.418, 0.601, 0.118, 0.626, 0.835, 0.361, 0.336.

c. Repeat b) for $\delta = 0.7, 0.5, 0.3, 0.2, 0.1$.
d. Observe the trend of the perturbed paths $\boldsymbol{X}_\delta$. In particular, when δ is small, most likely the perturbed parts from the jumping point to the merging point are the same as if they follow the original transition probabilities $p(j|i)$, $i, j = 1, 2, \ldots, \mathcal{S}$.

2.3. Let $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}$ be two independent ergodic Markov chains with the same transition probability matrix P on the same state space $\mathcal{S}$. Define $\boldsymbol{Y} = (\boldsymbol{X}, \widetilde{\boldsymbol{X}})$.

a. Prove that $\boldsymbol{Y}$ is ergodic.
b. Express L^*_{ij} in Figure 2.6 in terms of the Markov chain $\boldsymbol{Y}$.

2.4. Consider a three-state Markov chain with

$$P = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.1 & 0.6 & 0.3 \\ 0.7 & 0.1 & 0.2 \end{bmatrix}, \qquad f = \begin{bmatrix} 10 \\ 5 \\ 8 \end{bmatrix}.$$

a. Solve the Poisson equation (2.12) $(I - P)g + \eta e = f$ for g and η (by, e.g., setting $g(0) = 0$).
b. Solve $\pi = \pi P$ and $\pi e = 1$ for π first. Then, solve $(I - P + e\pi)g = f$ for g.
c. Compare both methods in a) and b).

2.5. For an ergodic Markov chain $\boldsymbol{X} = \{X_l, l = 0, 1, \ldots\}$, derive the Poisson equation using

$$g(i) = \lim_{L \to \infty} \sum_{l=0}^{L-1} E\{[f(X_l) - \eta] | X_0 = i\}.$$

2.6. The Poisson equation for the perturbed Markov chain is

$$(I - P_\delta)g_\delta + \eta_\delta e = f_\delta,$$

where $P_\delta = P + \delta\Delta P$ and $f_\delta = f + \delta\Delta f$. Derive the performance derivative formula (2.26) from the above equation.

2.7. Prove the following results:

a. If $f = ce$ with c being a constant, then $g = ce$ is a constant vector.
b. If $p(j|i) = p_j$ for all $i \in \mathcal{S}$; i.e., every row in the transition probability matrix is the same, then $g = f$.
c. If $p(j|i) = p(i|j)$, for all $i, j \in \mathcal{S}$; i.e., the transition probability matrix P is symmetric, then $\sum_{i=1}^{S} g(i) = \sum_{i=1}^{S} f(i)$.

2.8. Prove $e\frac{d\eta_\delta}{d\delta} = \lim_{\beta\uparrow 1} \frac{d\eta_{\beta,\delta}}{d\delta}$. In other words,

$$\frac{d}{d\delta}\left(\lim_{\beta\uparrow 1} \eta_{\beta,\delta}\right) = \lim_{\beta\uparrow 1} \frac{d\eta_{\beta,\delta}}{d\delta}.$$

2.9. Assume that P changes to $P_\delta = P + \delta(\varDelta P)$, $\varDelta Pe = 0$, and $f_\delta \equiv f$. Derive the second-order derivative of the discounted reward $\eta_{\beta,\delta}$ with respect to δ, $\frac{d^2\eta_{\beta,\delta}}{d\delta^2}$.

2.10. In Example 2.2, we have

$$G_1 := \varDelta P(I - P + e\pi)^{-1} = \begin{bmatrix} -3.2 & 3.2 \\ 3.2 & -3.2 \end{bmatrix}.$$

a. Find the eigenvalues and eigenvectors of G_1.
b. Verify that

$$\begin{bmatrix} -3.2 & 3.2 \\ 3.2 & -3.2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & -6.4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^{-1}.$$

c. Prove that

$$G_1^n = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & (-6.4)^n \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^{-1},$$

and

$$\begin{aligned} \pi_\delta &= \pi \sum_{n=0}^{\infty} G_\delta^n = \pi \sum_{n=0}^{\infty} (\delta G_1)^n \\ &= \pi \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & \sum_{n=0}^{\infty}(-6.4\delta)^n \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^{-1}. \end{aligned}$$

d. Determine the convergence region of π_δ. Extend the discussion to more general case.

2.11. A group is a nonempty set G, together with a binary operation on G, denoted as juxtaposition ab, $a, b \in G$, and $ab \in G$, with the following properties: (i) *(Associativity)* $(ab)c = a(bc)$, for all $a, b, c \in G$; (ii) *(Identity)* There exists an element $e \in G$ for which $ea = ae = a$ for all $a \in G$; and (iii) *(Inverse)* For each $a \in G$, there is an element denoted a^{-1}, for which $aa^{-1} = a^{-1}a = e$, [220].

a. Verify that the set of matrices defined in (2.50) with matrix multiplication as the juxtaposition satisfies the above properties.
b. In Example 2.2, we have

$$B = P - I = \begin{bmatrix} -0.10 & 0.10 \\ 0.15 & -0.15 \end{bmatrix}.$$

What is its group inverse? Is the inverse an infinitesimal generator?

2.12. Assume that the MacLaurin series of P_δ exists in $[0, \delta]$. Equation (2.57) can be derived directly by the following procedure: Taking the derivatives of the both sides of $\pi_\delta(I - P_\delta) = 0$ n times, we can obtain $\frac{d^n \pi}{d\delta^n}$ at $\delta = 0$. Then, we can construct the MacLaurin series of π. Work out the details of this approach and derive the MacLaurin series of η_δ at $\delta = 0$.

2.13. Prove the continuous version of the PRF equation (2.62) from its discrete version (2.7) by setting $B = P - I$, and vice versa.

2.14. Consider a Markov chain $\boldsymbol{X}$ with transition probabilities $p(j|i)$, $i, j \in \mathcal{S}$, and reward function f. For any $0 < p < 1$, we define an equivalent Markov chain $\boldsymbol{X}'$ with transition probabilities $p'(j|i) = (1 - p)p(j|i)$, $j \neq i$, and $p'(i|i) = p + (1 - p)p(i|i)$, $i \in \mathcal{S}$. Set $f' = f$. Prove that $\eta' = \eta$ and $g' = \frac{g}{1-p}$.

2.15. Consider a Markov process $\boldsymbol{X}$ with transition rates $\lambda(i)$, and transition probabilities $p(j|i)$, $i, j \in \mathcal{S}$, and reward function f. For any $\lambda > \lambda(i)$, $i \in \mathcal{S}$, we define an equivalent Markov process $\boldsymbol{X}'$ with transition rates $\lambda'(i) \equiv \lambda$, and transition probabilities $p'(j|i) = \frac{\lambda(i)}{\lambda} p(j|i)$, $j \neq i$, and $p'(i|i) = [1 - \frac{\lambda(i)}{\lambda}] + \frac{\lambda(i)}{\lambda} p(i|i)$. Set $f' = f$.

a. Prove that $\eta' = \eta$ and $g' = g$.
b. Let the discrete-time Markov chain embedded at the transition epoches of $\boldsymbol{X}'$ as $\boldsymbol{X}^\dagger$. Find the steady-state probability $\pi^\dagger$ and the potential $g^\dagger$ of $\boldsymbol{X}^\dagger$.
c. Suppose that $1 = \lambda > \lambda(i)$, $i \in \mathcal{S}$, prove that $g^\dagger = g$.
d. For any $\kappa > 0$, we define a Markov process $\widetilde{\boldsymbol{X}}$ with transition rates $\widetilde{\lambda}(i) = \kappa\lambda(i)$, $i \in \mathcal{S}$, transition probabilities $\widetilde{p}(j|i) = p(j|i)$, $i, j \in \mathcal{S}$, and reward function $\widetilde{f} = f$. Prove that $\widetilde{\pi} = \pi$ and $\widetilde{g} = \frac{g}{\kappa}$.
e. Given any Markov process $\boldsymbol{X}$, can you find a Markov chain that has the same steady-state probability π and potential g as $\boldsymbol{X}$? (*Hint: use the results in b)-d).)*

2.16.* For semi-Markov processes with the discounted reward defined in (2.93), set $\eta_\beta := (\eta_\beta(1), \ldots, \eta_\beta(S))^T$ and $g_\beta := (g_\beta(1), \ldots, g_\beta(S))^T$. Prove that (cf. [57])

$$\lim_{\beta \downarrow 0} g_\beta = g,$$

$$\lim_{\beta \downarrow 0} \eta_\beta = \eta e,$$

and

$$\eta_\beta = \beta g_\beta + \eta e.$$

2.17. Consider a two-server cyclic Jackson queueing network with service rates μ and λ for servers 1 and 2, respectively. There are N customers in the network. The system's state $\boldsymbol{n} = n$ is the number of customers at server 1. The state process is Markov. Let the performance be the average response time of the customers at server 1, denoted as $\bar{\tau}$. Calculate the performance potentials $g(i)$, $i = 1, 2, \ldots, S$, and the average response time $\bar{\tau}$, and derive the derivative of $\bar{\tau}$ with respect to λ and μ.

2.18. The two-server N-customer cyclic Jackson queueing network studied in Problem 2.17 is equivalent to an $M/M/1/N$ queue with arrival rate λ, service rate μ, and a finite buffer size N. (When the number of customers in the queue is $n = N$, an arriving customer is simply lost.)

a. Suppose that the arrival rate only changes when $n = 0$; i.e., when $n = 0$, λ changes to $\lambda + \Delta\lambda$, and when $n > 0$, λ remains unchanged. What is the derivative of the average response time $\bar{\tau}$ with respect to this change?
b. Suppose that the arrival rate only changes when $n = n^*$, with $0 < n^* < N$. What is the derivative of $\bar{\tau}$ with respect to this change?
c. Suppose that the arrival rate only changes when $n = N$. What is the derivative of $\bar{\tau}$ with respect to this change? (You may view the $M/M/1/N$ queue as the two-server cyclic queue again to verify your result.)

2.19. Consider a Markov chain with one closed recurrent state set $\mathcal{S}_1$ and one transient state set $\mathcal{S}_2$ (a uni-chain). Let the transition probability matrix be

$$P = \begin{bmatrix} P_1 & 0 \\ P_{21} & P_{22} \end{bmatrix},$$

with P_1 corresponding to $\mathcal{S}_1$ and P_{21}, P_{22} corresponding to $\mathcal{S}_2$, and 0 being a matrix with all zero components. Denote the potential vector as $g = (g_1^T, g_2^T)^T$ with $g_1 = (g(1), \ldots, g(S_1))^T$ and $g_2 = (g(S_1 + 1), \ldots, g(S))^T$, $S_1 = |\mathcal{S}_1|$, $S_2 = |\mathcal{S}_2|$, $S_1 + S_2 = S$.

Derive an equation for g_1 and express g_2 in terms of g_1 and P_{21}, P_{22}.

2.20. Consider a Markov chain with transition probability matrix

$$P = \begin{bmatrix} B & b \\ 0 & 1 \end{bmatrix},$$

where B is an $(S-1) \times (S-1)$ irreducible matrix, $b > 0$ is an $(S-1)$ dimensional column vector, 0 represents an $(S-1)$-dimensional row vector whose components are all zero. The last state S is an absorbing state. Clearly, the long-run average reward for this Markov chain is $\eta = f(S)$, independent of B, b, and the initial state. Thus, the long-run average reward does not reflect the transient behavior. Now, we set $f(S) = 0$. Define

$$g(i) = E\left[\sum_{l=0}^{\infty} f(X_l) \Big| X_0 = i\right].$$

Let $L_{i,S} = \min\{l : l \geq 0, X_l = S | X_0 = i\}$ be the first passage time from i to S. Then,

$$g(i) = E\left[\sum_{l=0}^{L_{i,S}-1} f(X_l)\Big| X_0 = i\right].$$

a. Derive an equation for $g = (g(1), \ldots, g(S))^T$.
b. Derive an equation for the average first passage times $E[L_{i,S}]$, $i \in \mathcal{S}$.

2.21.* (This problem helps in understanding the difference between the discounted reward criteria for both the discrete-time and continuous-time models.) Consider a Markov chain $\boldsymbol{X}$ with transition probability matrix $P = [p(j|i)]_{i,j=1}^{S}$ and reward function $f(i)$, $i = 1, 2, \ldots, S$. For simplicity, we assume that $p(i|i) = 0$ for all $i = 1, 2, \ldots, S$. Let $\widetilde{\boldsymbol{X}}$ be a Markov chain with reward function $\widetilde{f}(i) = f(i)$, $i = 1, 2, \ldots, S$, and transition probability matrix $\widetilde{P}$ defined as $\widetilde{p}(i|i) = q$, $0 < q < 1$, and $\widetilde{p}(j|i) = (1-q)p(j|i)$, $j \neq i$, $i, j = 1, 2, \ldots, S$.

a. Prove that $\widetilde{\boldsymbol{X}}$ is equivalent to $\boldsymbol{X}$ in the sense that they have the same steady-state probabilities: $\widetilde{\pi}(i) = \pi(i)$ for all $i = 1, 2, \ldots, S$.
b. The discounted reward of $\boldsymbol{X}$ is defined as (2.30):

$$\eta_\beta(i) = (1-\beta)E\left[\sum_{l=0}^{\infty} \beta^l f(X_l)\Big| X_0 = i\right],$$

where $0 < \beta < 1$ is a discount factor. Similarly, the discounted reward of $\widetilde{\boldsymbol{X}}$ is defined with a discount factor $0 < \widetilde{\beta} < 1$ as

$$\widetilde{\eta}_{\widetilde{\beta}}(i) = (1-\widetilde{\beta})E\left[\sum_{l=0}^{\infty} \widetilde{\beta}^l f(\widetilde{X}_l)\Big| \widetilde{X}_0 = i\right].$$

Find a value for $\widetilde{\beta}$ such that $\widetilde{\eta}_{\widetilde{\beta}}(i) = \eta_\beta(i)$ for all $i = 1, 2, \ldots, S$.
c. Let $\Delta > 0$ be a positive number. Consider a continuous-time (non-Markov) process $\hat{\boldsymbol{X}} := \{\hat{X}_t, t \in [0, \infty)\}$, where $\hat{X}_t = X_l$ if $l\Delta \le t < (l+1)\Delta$, $l = 0, 1, \ldots$, with $\boldsymbol{X} = \{X_l, l = 0, 1, \ldots\}$ being the Markov chain considered in a). The discounted reward of $\hat{\boldsymbol{X}}$ is defined by an exponential weighting factor (cf. (2.93)):

$$\eta_\alpha(i) = \lim_{T\to\infty} E\left[\int_0^T \alpha \exp(-\alpha t) f(\hat{X}_t) dt\Big| X_0 = i\right], \qquad T_0 = 0.$$

What is the equivalent β such that $\eta_\beta(i) = \eta_\alpha(i)$ for all $i = 1, 2, \ldots, S$?
d. Repeat c) for continuous-time process $\hat{\boldsymbol{X}} := \{\hat{X}_t, t \in [0, \infty)\}$, with $\hat{X}_t = \widetilde{X}_l$ if $l\Delta \le t < (l+1)\Delta$, $l = 0, 1, \ldots$.
e. What about in d) when we let $\Delta \to 0$ while keeping $\frac{1-q}{\Delta} = \lambda$ (where λ is a constant)?

(Hint: If $\boldsymbol{X} = \{X_0 = i_0, X_1 = i_1, \ldots\}$, then we have $\widetilde{\boldsymbol{X}} = \{\widetilde{X}_0 = \widetilde{X}_1 = \cdots = \widetilde{X}_{n_0-1} = i_0, \widetilde{X}_{n_0} = \widetilde{X}_{n_0+1} = \cdots = \widetilde{X}_{n_0+n_1-1} = i_1, \ldots\}$, where n_l is the number of consecutive visits to state i_l, $l = 0, 1, \ldots$. Note that n_l is geometrically distributed with parameter q. Therefore,

$$\widetilde{\eta}_{\widetilde{\beta}}(i) = (1-\widetilde{\beta})E[(1+\widetilde{\beta}+\cdots+\widetilde{\beta}^{n_0-1})f(i_0)+(\widetilde{\beta}^{n_0}+\cdots+\widetilde{\beta}^{n_0+n_1-1})f(i_1)+\cdots].$$

We conclude that $\widetilde{\eta}_{\widetilde{\beta}}(i) = \eta_\beta(i)$ if $\beta = \frac{(1-q)\widetilde{\beta}}{1-q\widetilde{\beta}}$.)

2.22. Prove that the random variable s generated according to (2.96) is indeed exponentially distributed.

2.23. Develop a PA algorithm to determine a perturbed sample path for an open Jackson network consisting of M servers, with mean service time $\bar{s}_i$, $i = 1, 2, \ldots, M$. The customers arrive in a Poisson process with mean inter-arrival time $a = \frac{1}{\lambda}$. Both a and $\bar{s}_i$, $i = 1, 2, \ldots, M$, may be perturbed.

2.24. Suppose that at some time the perturbations of the servers in a closed network are $\Delta_1, \Delta_2, \ldots, \Delta_M$ determined by Algorithm 2.1. What is the perturbation that has been realized by the network at that time? As we know, if a perturbation is realized, then the future perturbed sample path looks the same as the original one except that it is shifted to the right by an amount equal to the perturbation. Can we use this fact to simplify the calculation in Algorithm 2.2?

2.25. Using the 0-1 vector array (2.105), discuss the situation of the propagation of M perturbations with the same size, each at one server, along a sample path. Prove that $\sum_{i=1}^{M} c(\boldsymbol{n}, i) = 1$.

2.26. We further study the propagations of two equal perturbations $\Delta_1 = \Delta$ at server 1 and $\Delta_2 = \Delta$ at server 2 simultaneously on the same sample path. Consider the array in (2.105). Set $w(t) = w_1(t) + w_2(t)$.

a. What is the meaning of $w(t)$?
b. What does it mean when $w(t) = (1, 1, \ldots, 1)$ or $w(t) = (0, 0, \ldots, 0)$?
c. How does $w(t)$ evolve?

2.27. In addition to (2.94), we may define the system performance as the long-run time average

$$\eta_T^{(f)} = \lim_{L\to\infty} \frac{1}{T_L} \int_0^{T_L} f[\boldsymbol{N}(t)]dt.$$

We have $\eta_T^{(f)} = \frac{\eta^{(f)}}{\eta^{(I)}}$.

a. Derive the derivative of $\eta_T^{(f)}$ with respect to $\bar{s}_i$, $i = 1, 2, \ldots, M$.

b. Define the reward function f corresponding to the steady-state probability $\pi(\boldsymbol{n})$, with $\boldsymbol{n}$ being any state, and derive $\frac{d\pi(\boldsymbol{n})}{d\bar{s}_i}$, $i = 1, 2, \ldots, M$.

2.28.* Prove that, in a closed Jackson network, the sample function $T_L(\xi, \bar{s}_v)$ (with ξ fixed) is a piecewise linear function of $\bar{s}_v$, $v = 1, 2, \ldots, M$ (see [46]).

2.29. Consider a closed Jackson network in which $\mu_i q_{i,j} = \mu_j q_{j,i}$, $i, j = 1, 2, \ldots, M$. Prove that

$$c(\boldsymbol{n}, k) = \frac{n_k}{N}, \qquad k = 1, 2, \ldots, M;$$

and

$$\frac{\bar{s}_k}{\eta} \frac{\partial \eta}{\partial \bar{s}_k} = -\frac{1}{M},$$

where $k = 1, 2, \ldots, M$, denotes any server in the network.

2.30.* *(This problem requires a good knowledge of queueing theory)* Consider an M/M/1 queue with arrival rate λ and service rate μ. The system state is simply the number of customers in the queue; i.e., $\boldsymbol{n} = n$. The performance measure is the average response time $\tau = \lim_{L\to\infty} \frac{1}{L} \int_0^{T_L} n(t)dt$. Thus, $f(n) = n$. For the M/M/1 queue, there is a source sending customers to the queue with rate λ. Denote the source as server 0, and the server as server 1. Server 0 can be viewed as always having infinitely many customers.

a. Prove that the realization factors $c^{(f)}(n, 0)$ and $c^{(f)}(n, 1)$, $n = 0, 1, \ldots,$ satisfy the following equations:

$$c^{(f)}(0, 0) = 0, \; c^{(f)}(0, 1) = 0,$$

$$c^{(f)}(n, 0) + c^{(f)}(n, 1) = n, \qquad n \geq 0,$$

$$(\lambda + \mu)c^{(f)}(n, 0) = \mu c^{(f)}(n - 1, 0) + \lambda c^{(f)}(n + 1, 0) - \lambda, \qquad n > 0,$$

and

$$(\lambda + \mu)c^{(f)}(n, 1) = \lambda c^{(f)}(n + 1, 1) + \mu c^{(f)}(n - 1, 1) + \mu, \qquad n > 0.$$

b. To solve for $c^{(f)}(n, i)$, $i = 0, 1$, we need a boundary condition. Using the physical meaning of perturbation realization, prove that $c^{(f)}(1, 1)$ equals the average number of customers served in a busy period of the M/M/1 queue; i.e. (see, e.g., [169]),

$$c^{(f)}(1, 1) = \frac{\mu}{\mu - \lambda} = \frac{1}{1 - \rho}, \qquad \rho = \frac{\lambda}{\mu}.$$

c. Prove

$$c^{(f)}(n, 1) = \frac{n}{1 - \rho},$$

and

$$c^{(f)}(n, 0) = -\frac{n\rho}{1 - \rho}.$$

d. By the same argument as in closed networks, explain and derive

$$\frac{\mu}{\eta^{(I)}}\frac{d\tau}{d\mu} = -\frac{\lambda\mu}{(\mu-\lambda)^2} = -\frac{\rho}{(1-\rho)^2},$$

and

$$\frac{\lambda}{\eta^{(I)}}\frac{d\tau}{d\lambda} = \frac{\lambda^2}{(\mu-\lambda)^2} = \frac{\rho^2}{(1-\rho)^2}.$$

2.31. The head-processing time of a packet in a communication system, or the machine tool set-up time in manufacturing, is usually a fixed amount of time. Consider a two-server cyclic queueing network in which the service times of the two servers are exponentially distributed with mean $\bar{s}_1$ and $\bar{s}_2$, respectively. Suppose that every service time of server 1 increases by a fixed amount of time Δ. Derive the derivative of performance $\eta^{(f)}$ with respect to Δ using performance realization factors $c^{(f)}(\boldsymbol{n}, 1)$.

2.32. Prove that Algorithm 2.2 yields a strongly consistent estimate for the derivative of the average response time in an M/G/1 queue; i.e., in (2.111) we have

$$\frac{\mu}{\eta^{(I)}}\frac{\partial\bar{\tau}}{\partial\mu} = -\lim_{K\to\infty}\frac{1}{T_L}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\sum_{l=1}^{i} s_{k,l}, \qquad \text{w.p.1.}$$

2.33. Consider a closed Jackson network with M servers and N customers. The throughput of server i is $\eta_i = \breve{\eta} v_i$ where $\breve{\eta}$ is the "un-normalized system throughput":

$$\breve{\eta} = \frac{G_M(N-1)}{G_M(N)},$$

where v_i is server i's visiting ratio: The solution to

$$v_i = \sum_{j=1}^{M} q_{j,i} v_j, \qquad j = 1, 2, \ldots, M,$$

and (see (C.16) in Appendix C)

$$G_m(n) = \sum_{n_1+\ldots+n_M=n}\prod_{i=1}^{m} x_i^{n_i},$$

where $x_i = v_i\bar{s}_i$, $i = 1, 2\ldots, M$. We have

$$dx_i = dv_i\bar{s}_i + v_i d\bar{s}_i. \tag{2.128}$$

Now, we consider the derivative of $\breve{\eta}$ with respect to the routing probability matrix $Q = [q_{i,j}]_{i,j=1}^{M}$. It is clear that $\breve{\eta}$ depends on the routing probabilities only through x_i, $i = 1, 2, \ldots, M$. Suppose that v_i changes to $v_i + dv_i$, $i = 1, 2, \ldots, M$. From (2.128), we observe that in terms of the changes in x_i, dx_i, $i = 1, 2, \ldots, M$, this is equivalent to setting $dv_i = 0$ and $d\bar{s}_i = \bar{s}_i\frac{dv_i}{v_i}$ for all $i = 1, 2, \ldots, M$.

a. Explain that, for closed Jackson networks, the derivative of the average reward $\sum_{\text{all } \boldsymbol{n}} \pi(\boldsymbol{n}) f(\boldsymbol{n})$ with respect to the changes in routing probabilities can be obtained through the derivatives of the average reward with respect to the mean service times.
b. Derive the performance derivative formula $\frac{d\eta_i}{dQ}$, by using performance realization factors $c^{(f)}(\boldsymbol{n}, i)$, $i = 1, 2, \ldots$.

2.34.* Consider the same two-server cyclic Jackson queueing network studied in Problem 2.17. Let $\eta_T^{(f)} = \lim_{L\to\infty} \frac{\int_0^{T_L} f(n(t))dt}{T_L}$ denote the time-average performance, where $n(t)$ is the number of customers at time t at server 1, and L denotes the number of transitions. The performance function is $f(n) = n$. Let us assume that the arrival rate λ, or the service rate μ, changes only when the state is n.

a. Derive $\frac{d\eta_T^{(f)}}{d\lambda}$ and $\frac{d\eta_T^{(f)}}{d\mu}$ in terms of the realization factors $c^{(f)}(n, 1), c^{(f)}(n, 2)$ and realization probability $c(n, 1), c(n, 2)$.
b. Express $\frac{d\eta_T^{(f)}}{d\lambda}$ and $\frac{d\eta_T^{(f)}}{d\mu}$ in terms of the performance potentials $g(n)$.
c. Compare both results in a) and b) and derive a relation between the realization factors and the potentials. Give an intuitive explanation for this relation. (cf. [260])

2.35. In weak derivative expression (2.125), we may choose $P^+ = P'$ and $P^- = P$.

a. Derive (2.126) and express its meaning based on sample paths.
b. Derive (2.127).

2.36. Derive (2.23) from (2.127).

2.37. Consider a (continuous-time) Markov process with transition rates $\lambda(i)$ and transition probabilities $p(j|i)$, $i, j = 1, 2, \ldots, S$. Suppose that the transition probability matrix $P := [p(j|i]_{i,j\in\mathcal{S}}$ changes to $P+\delta\Delta P$ and the transition rates $\lambda(i)$, $i = 1, 2, \ldots, S$ remain unchanged. Let η be the average reward with reward function f. Derive the performance derivative formula for $\frac{d\eta_\delta}{d\delta}$ using the construction approach illustrated in Section 2.1.3.